...ICAL ROOM
...ERAL LIBRARY
...SITY OF MICH.

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

HAROLD RUGG, *Editor*

VOL. XXII, No. 9 — DECEMBER, 1931 — $5.00 per year

CONTENTS

Title Page and Index to Volume XXII (1931) . . . . . i–viii

THE ROUTINE COMPUTATION OF PARTIAL AND MULTIPLE CORRELATION. *Raymond Franzen and Mahew Derryberry* 641

A STUDY OF QUESTIONNAIRE TECHNIQUE. *Frank K. Shuttleworth* 652

INFLUENCE OF THE ASSIGNMENT ON LEARNING. *David H. Briggs and A. M. Jordan* 659

FACTUAL MEMORY OF SECONDARY SCHOOL PUPILS FOR A SHORT ARTICLE WHICH THEY READ A SINGLE TIME. (*Concluded from November Issue*) *Alfred G. Dietze and George Ellis Jones* 667

EFFECT OF ORDER OF PRESENTATION ON THE RECALL OF PICTURES. *Daniel D. Droba* 677

THE INFLUENCE OF AN AUDIENCE UPON RECALL. *Clara Burri* . 683

THE RELATIVE INFLUENCE OF VISUAL AND AUDITORY FACTORS IN SPELLING ABILITY. *George W. Hartmann* 691

FURTHER DATA CONCERNING THE EFFECT OF WEIGHTING EXERCISES IN NEW-TYPE EXAMINATIONS. *C. W. Odell* 700

THE HANDWRITING OF INDIANS. *Thomas R. Garth* 705

A MEASUREMENT OF THE KNOWLEDGE OF PSYCHOLOGY BEFORE AND AFTER FORMAL TRAINING. *Calvin Hall* 710

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — **Baltimore, Md.**

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

ESTABLISHED 1910

Editor: HAROLD RUGG

Board of Associate Editors

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — E. L. THORNDIKE
FRANK N. FREEMAN — V. A. C. HENMON — L. M. TERMAN — GOODWIN B. WATSON

Assistant to the Editor: FRANCES M. FOSTER

The Journal is published monthly, except June to August, by Warwick and York, Inc., Baltimore, Md., and York, Pa.

The price is $5.00 a year, payable in advance; foreign postage 30 cents extra. Single current issues are 70 cents each, and subscriptions for less than a year are charged at single issue rate for each number ordered. Back volumes are $6.00 each; single back numbers, 90 cents each, except, when more than three years old, the price is $1.00 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Manuscripts for publication, books or other materials for review, and correspondence should be addressed to the publishers.

WARWICK AND YORK, INC., *Baltimore, Maryland*

TEACHERS' HEADQUARTERS

European Plan

Entirely Fire-Proof

Centrally Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

The Psychology of Learning
1928 Revision — BY WILLIAM H. PYLE
$2.20+12¢ postage
Warwick & York, Inc., *Baltimore, Maryland*

An Experimental Study of Certain Factors Affecting

Transfer of Training in Arithmetic

By JAMES ROBERT OVERMAN

The investigation here reported was planned to study the effect of instruction on three types of examples in two-place addition upon the pupils' ability to handle closely related types in addition and subtraction, and to determine whether the amount of transfer is a function of the method of teaching.

The problem was to ascertain if the amount of such transfer can be increased by helping the pupils consciously to generalize the process and to formulate, from the types taught, a general method of procedure applicable to related types.

An introductory chapter summarizes briefly previous experiments in the transfer of training and shows the changes that have occurred in the attitude of psychologists toward this question within the last thirty-five years.

$3.00 plus 12 cents postage

WARWICK & YORK, INC.
Publishers
BALTIMORE, MD.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXII | December, 1931 | Number 9

THE ROUTINE COMPUTATION OF PARTIAL AND MULTIPLE CORRELATION

RAYMOND FRANZEN AND MAHEW DERRYBERRY

American Child Health Association, New York City

Probably the simplest and most direct method of computing multiple correlation and the regression coefficients of a multiple regression equation is the formation of a system of normal equations from the moments and product moments and the solution of these equations by the Doolittle method as reported by Tolley and Ezekiel in Vol. XVIII of the *Journal of the American Statistical Association.* A modified form of this method to be used when the intercorrelations are known and the raw moments are not available was published by Garrett.* It often happens, however, that not only are the regression coefficients and the multiple correlation desired, but the partial correlations are needed for interpretative purposes. In such problems it is more advantageous to compute the multiple correlation from the partial correlations according to the usual formula:

$$R_{1.234\ldots n} = \sqrt{1 - (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23}) \cdots (1 - r^2_{1n.23\ldots n-1})}$$

and the regression coefficients from the formula:

$$b_{12.34\ldots n} = r_{12.34\ldots n} \frac{\sigma_1 \sqrt{(1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23}) \cdots (1 - r^2_{1n.23\ldots n-1})}}{\sigma_2 \sqrt{(1 - r^2_{12})(1 - r^2_{23.1})(1 - r^2_{24.13}) \cdots (1 - r^2_{2n.13\ldots n-1})}}$$

But the computation of $r_{1n.234\ldots n-1}$ and the other higher order partials becomes increasingly laborious as the number of secondary subscripts are increased. It can be shown that the formula

$$n + 2(n-1) + 3(n-2) + \cdots n(n - \overline{n-1}) \qquad (1)$$

* A Modification of Tolley and Ezekiel's Method of Handling Multiple Correlation Problems. *Journal of Educational Psychology*, January, 1928.

gives the minimum number of partials that are necessary for computing a partial of the *n*th order; or, in other words, it gives the minimum number of times the usual formula

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{23}}}$$

must be applied to finally obtain a partial of the *n*th order. It will be noted that formula (1) gives the *minimum* number of computations. The possible number that may be computed is many more than that given by formula (1). For example, to compute a fourth order partial it is possible to compute thirty-five separate partials before finally obtaining a fourth order partial. On the other hand, only twenty partials need be computed if the most economical combinations of correlations are made in the computation of the lower order coefficients. Even though the computations are made in such a way that the minimum number of partials are computed, the usual method still requires considerable attention to avoid an incorrect combination in one of the lower order partials which would make all the subsequent work erroneous.

The method here presented completely eliminates these difficulties by reducing the solution to mechanical routine. It claims the following advantages:

1. The *n*th order partial correlation is obtained in the minimum number of operations, since the possibility of duplication of computations is eliminated.

2. Correct combinations of lower order partials are routinely made assuring the accurate higher order partial.

3. All computations are automatically checked without a double computation as is often advised.

4. Repeated writing of figures is eliminated.

5. Because of the direct mechanical and systematic calculation all the computations from first order correlations to the multiple can be done by a clerk.

6. A compact, concise record of the computations is provided for filing so that all partials of a given order and denomination are kept on the same page.

7. The multiple correlation is obtained from the partial forms with a minimum amount of computation.

The explanation below may seem difficult and complicated, but the routine is simple and easily followed once the method is under-

DIAGRAM I

(1)	r_{12}			
(2)	$r_{15} \cdot r_{25}$			
(3)	$r_{12} - r_{15}r_{25}$			
(4)	$k_{15} \cdot k_{25}$			
(5)	$r_{12.5}$			
(6)	r_{13}	r_{23}		
(7)	$r_{15} \cdot r_{35}$	$r_{25} \cdot r_{35}$		
(8)	$r_{13} - r_{15}r_{35}$	$r_{23} - r_{25}r_{35}$		
(9)	$k_{15} \cdot k_{35}$	$k_{25} \cdot k_{35}$		
(10)	$r_{13.5}$	$r_{23.5}$		
(11)	r_{14}	r_{24}	r_{34}	
(12)	$r_{15} \cdot r_{45}$	$r_{25} \cdot r_{45}$	$r_{35} \cdot r_{45}$	
(13)	$r_{14} - r_{15}r_{45}$	$r_{24} - r_{25}r_{45}$	$r_{34} - r_{35}r_{45}$	
(14)	$k_{15} \cdot k_{45}$	$k_{25} \cdot k_{45}$	$k_{35} \cdot k_{45}$	
(15)	$r_{14.5}$	$r_{24.5}$	$r_{34.5}$	
(16)	r_{15}	r_{25}	r_{35}	r_{45}
(17)	k_{15}	k_{25}	k_{35}	k_{45}
(18)				
(19)		(Check 1) $r_{45} + r_{35} + r_{25}$	(Check 2) $r_{45} + r_{35}$	(Check 3) r_{45}
(20)		(Check 4) $k_{45} + k_{35} + k_{25}$	(Check 5) $k_{35} + k_{45}$	(Check 6) k_{45}
(21)	$r_{12} + r_{13} + r_{14}$	$r_{23} + r_{24}$	r_{34}	
(22)	(r_{15}) (Check 1)	(r_{25}) (Check 2)	(r_{35}) (Check 3)	
(23)	Σ Numerator	Σ Numerator	Σ Numerator	
(24)	(k_{15}) (Check 4)	(k_{25}) (Check 5)	(k_{35}) (Check 6)	

DIAGRAM II

$r_{12.5}$		
$r_{13.5}$	$r_{23.5}$	
$r_{14.5}$	$r_{24.5}$	$r_{34.5}$

stood. The computational form with the entries and checks in terms of symbols is given in Diagram I. Here the operations necessary to compute a set of first order partials from the interrelationship of five variables are indicated. Diagram II denotes the method of transferring these first order partials to a second form for the computation of second order partials.

To make the diagrams more concrete the multiple correlation of four bony measurements with weight and two regression coefficients have been computed from the table of intercorrelations given below, and each step in the computation explained. The use of Diagrams I and II with the explanation will reduce the seeming complexity of the detailed directions that follow.

TABLE I.—INTERCORRELATIONS OF ANTHROPOMETRIC TRAITS IN ONE THOUSAND ELEVEN-YEAR-OLD GIRLS*

	Weight	Height	Chest breadth	Chest depth
Height	.7428			
Chest breadth	.8337	.6298		
Chest depth	.8201	.5388	.6845	
Hip width	.8752	.7334	.7323	.6956

* Data gathered in connection with the School Health Study conducted by the American Child Health Association where this routine was devised and used.

STEPS IN THE COMPUTATION

1. To compute first order partials the r's are entered on specially ruled sheets (see Diagram III) just as they would be entered in an inter-r table with the exception that four lines are left below each r for the entry of computational figures. The order in which the coefficients are entered determines the order in which the variables will be partialled out, and before beginning computation this order must be decided upon so that the r's will be so arranged on the sheet as to take care of this demand. The correlations of all other traits with the variable to be partialled out first are entered as the last row of intercorrelations, the correlations of all other traits with the variable to be partialled out second is placed second from the bottom, etc. Thus in this computation we have chosen to partial out hip width from all the remaining inter-r's first and chest breadth second. (Compare Table I and Diagram III noticing arrangement of variables.)

2. Immediately below each of the correlations in the last line (16) of intercorrelations (the correlations with the variable to be partialled out first) is written the k or $\sqrt{1-r^2}$ of the coefficient appearing directly above it (line 17). These values of $\sqrt{1-r^2}$ are taken from "Tables of $\sqrt{1-r^2}$ and $1-r^2$" by Miner.

DIAGRAM III*

		I Weight	II Height	III Chest Depth	IV Chest Breadth
Height	(1)	.7428			
	(2)				
	(3)				
	(4)				
	(5)				
Chest Depth	(6)	.8201	.5388		
	(7)				
	(8)				
	(9)				
	(10)				
Chest Breadth	(11)	.8337	.6298	.6845	
	(12)				
	(13)				
	(14)				
	(15)				
Hip Width	(16)	.8752	.7334	.6956	.7323
k	(17)	.483761	.679797	.718429	.680982
	(18)				
	(19)		2.1613	1.4279	.7323
	(20)		2.079208	1.399411	.680982
	(21)	2.3966	1.1686	.6845	
	(22)				
	(23)				

* To simplify the explanation in the text the lines and columns are numbered. The numbers are not necessary in the ordinary computational procedure.

The computations described in Steps 3, 4 and 5 below are used in the check and are not necessary for the computation of the partials.

3. Beginning at the extreme right of the last row of intercorrelations (line 16) add each figure towards the left recording each cumulative total three lines below (line 19), and in the column corresponding to the last figure used in the cumulative total. Thus, .7323 is written

in the first column; .7323 + .6956 is written in the second column, and 1.4279 + .7334 is written in the third, and so on towards the left until all of the r's have been added except the r in the last column to the left. It is not necessary to include this r in the summation.

4. The next line (line 20) is obtained from the line of k's (line 17) beginning similarly with the figure at the extreme right, and adding each figure in succession with the exception of the k in the last column to the left.

5. Sum each column of inter-r's omitting the last line of r's (line 16) and record on line 21. For example, 2.3966 is the sum of .7428 + .8201 + .8337.

Computation steps 6 to 14 are illustrated on Diagram IV.

6. If a calculating machine is used put the r that is in column I and in the bottom row (.8752 in row 16) in the machine and multiply by each succeeding r in that row. The products are written beneath each of the zero order r's in column I in succession. Thus the .6419 entered on line 2 is .8752 × .7334; .6088 entered on line 7 under the next zero r in the first column is the product of .8752 × .6956 and .6409 on line 12 is .8752 × .7323.

7. Similarly put the r in column II and in the bottom row (.7334 in row 16) in the machine and multiply by each succeeding r in that row and record the products in column II in succession on lines 7 and 12. The entry in column III, line 12, is the product of the r in column III on the last line (.6956) and the r in column IV (.7323). These computations give the encircled term of the formula below for each of the six partials on Diagram IV.

$$r_{xy.z} = \frac{r_{xy} - \boxed{r_{xz}r_{zy}}}{k_{xz}k_{zy}}$$

8. Subtract each of these products obtained in Step 7 from the r just above it in the column. Thus, .1009 in line 3 is .7428 − .6419. (The operations indicated are, of course, always performed algebraically, giving proper regard for signs.) This is the numerator of the formula given above.

To check result thus far:

9. Multiply the r in the bottom row of column I (line 16) by the figure in column II, line 19, and enter the product in column I, line 22. Thus, 1.8916 is the product of .8752 × 2.1613. Likewise multiply the r in the last row of column II by the figure in column III, line 19, and record the product in column II, line 22; *i.e.*, the 1.0473 is the

product of .7334 × 1.4279. The remaining product in column III, line 22, is found in a similar manner.

10. Subtract the products obtained in Step 9 from the summations immediately above them on line 21, and enter the differences in line 23.

11. Sum the numerators in column I on lines 3, 8 and 13 obtained in Step 8. The sum should equal the figure shown in column I, line 23, as the result of Step 10, (*i.e.* .1009 + .2113 + .1928 = .5050).* Of course in some cases there may be a difference of 1 or 2 points in the last decimal place due to dropping of decimals. In a similar manner columns II and III are checked.

To compute the denominators:

12. Perform the multiplications described in Step 6 using the row of *k*'s, line 17, instead of *r*'s, line 16, and enter the products on the lines below the several numerators obtained in Step 8. This gives the denominator of the formula above; (*i.e.* .3289 = .483761 × .679797).

To check these entries:

13. Multiply the *k* in column I by the summation in column II on line 20. Record in column I, line 24. Sum the denominators in column I obtained from Step 12, lines 4, 9 and 14. The two results should check. Similarly multiply the *k* in column II by the summation in column III and the product should equal the sum of the products of *k's* in column II.

Final computations:

14. Divide the numerators obtained in Step 8 by the respective denominators obtained in Step 12. The results are usually written in red (italics here) or otherwise made distinctive so they can be picked out readily. This quotient is the partial. For example, .1009 ÷ .3289 gives .3068, which is the partial correlation of height and weight independent of the influence of width of hips. Similarly .3579 is the relation of chest breadth to chest depth independent of hip width. The only check on this division is the multiplication of the denominator by the quotient, obtaining the numerator.

To compute second order partials:

15. The first order partials obtained on Diagram IV are entered as shown in Diagram V. The arrangement is the same as Diagram III except that no entries are made on line 16, since all the correlations

---

* This may be shown to be true by the following equation:

$$(r_{12} - r_{15}r_{25}) + (r_{13} - r_{15}r_{35}) + (r_{14} - r_{15}r_{45}) = r_{12} + r_{13} + r_{14} - r_{15}(r_{25} + r_{35} + r_{45})$$

with hip width have been partialled out. The last line is now chest breadth. These entries are checked by summations of the successive columns of Diagram IV agreeing with the summations of the corresponding column of Diagram V. By performing the computations described in Steps 2 to 14, the variable chest breadth will be partialled

DIAGRAM IV*

		I Weight	II Height	III Chest Depth	IV Chest Breadth
Height	(1)	.7428			
	(2)	.6419			
	(3)	.1009			
	(4)	.3289			
	(5)	*.3068*			
Chest *D*	(6)	.8201	.5388		
	(7)	.6088	.5102		
	(8)	.2113	.0286		
	(9)	.3475	.4884		
	(10)	*.6081*	*.0586*		
Chest *B*	(11)	.8337	.6298	.6845	
	(12)	.6409	.5371	.5094	
	(13)	.1928	.0927	.1751	
	(14)	.3294	.4629	.4892	
	(15)	*.5853*	*.2003*	*.3579*	
Hip Width *k*	(16)	.8752	.7334	.6956	.7323
	(17)	.483761	.679797	.718429	.680982
	(18)				
	(19)		2.1613	1.4279	.7323
	(20)		2.079208	1.399411	.680982
	(21)	2.3966	1.1686	.6845	
	(22)	1.8916	1.0473	.5094	
	(23)	.5050	.1213	.1751	
	(24)	1.0058	.9513	.4892	

* Diagram IV is like Diagram III with the remaining computations completed.

out. The results obtained from these computations are all relationships of two of the variables irrespective of hip width and chest breadth.

Each of the successive higher order partials can be obtained by repeating the process, partialling out one additional variable with each repetition.

DIAGRAM V

	I Weight	II Height	III Chest Depth
Height	(1) .3068		
	(2) .1172		
	(3) .1896		
	(4) .7944		
	(5) *.2387*		
Chest *D*	(6) .6081	.0586	
	(7) .2094	.0717	
	(8) .3986	−.0131	
	(9) .7571	.9148	
	(10) *.5265*	*−.0143*	
Chest *B*	(11) .5853	.2003	.3579
	(12) .810817	.979735	.933760
	(13)		
	(14)	.5582	.3579
	(15)	1.913495	.933760
	(16) .9149	.0586	
	(17) .3267	.0717	
	(18) .5882	−.0131	
	(19) 1.5515	.9148	

	Computation of 3rd Order Partials					
	(*a*)			(*b*)		
	I Weight			I Weight		
Height	.2387		Chest *D*	.5265		
	−.0075			−.0034		
	.2462			.5299		
	.8501			.9710		
	.2896	II Height		*.5457*	II Height	
Chest *D*	.5265	−.0143	Height	.2387	−.0143	
	.850175	.999898		.971093	.999898	

Computation of the second and third order partials are pictured in Diagram V. The third order partial (*a*) is obtained in the usual manner. The third order partial (*b*) is also given in this diagram and will be used later in the computation of the regression coefficients. It is obtained from a rearrangement of the second order correlations. This is always possible when computing partials of the order n–2 (where n is the number of variables involved), for

$$r_{12.34\ldots n} = \frac{r_{12.45\ldots n} - (r_{13.45\ldots n})(r_{23.45\ldots n})}{(k_{13.45\ldots n})(k_{23.45\ldots n})} \text{ and}$$

$$r_{13.24\ldots n} = \frac{r_{13.45\ldots n} - (r_{12.45\ldots n})(r_{23.45\ldots n})}{(k_{12.45\ldots n})(k_{23.45\ldots n})}$$

To compute the multiple correlation:

The multiple correlation coefficient may be expressed as a function of the alienation coefficients of the partial correlation, thus:

$$R_{1.234\ldots n} = \sqrt{1 - (k_{12})(k_{13.2})(k_{14.23}) \cdots (k_{1n.23\ldots n-1})}$$

If the criterion variable has been entered in column I in the computation of the successive higher order partials all of the k's required in the above formula have been determined except $k_{1n.23\ldots n-1}$. But $r_{1n.23\ldots n-1}$ has been computed so the corresponding k can be quickly determined from Miner's Tables. In the illustrative table the k of .2896 is .957148. By a multiplication of the k's in the first column on each successive sheet, and the k of the highest order partial in column I the multiple k or ($k_{1.234\ldots n}$) will be obtained. Thus $k_{1.2345}$ = .483761 (Diagram IV, line 17) × .810817 (Diagram V, line 12) × .850175 (Diagram V, line 7) × .957148 (k of .2896) = .3192. From this product $R_{1.2345}$ which, of course, is $\sqrt{1 - k_{1.2345}}$ is obtained from Miner's Tables. In this example $R_{1.2345}$ = .9477.

To compute the regression coefficient:

If we substitute in the multiple regression equation

$$\bar{X}_1 = b_{12.345}X_2 + b_{13.245}X_3 + b_{14.235}X_4 + b_{15.234}X_5 - C \qquad \text{(A)}$$

for $b_{12.345}$, $b_{13.245}$, etc. the partial r's and the partial σ's that are their equivalents the equation reduces to

$$\begin{aligned}\bar{X}_1 = {} & r_{12.345}\frac{\sigma_1}{\sigma_2}\frac{k_{1.2345}}{(k_{25})(k_{24.5})(k_{23.45})(k_{12.345})}X_2 \\ & + r_{13.245}\frac{\sigma_1}{\sigma_3}\frac{k_{1.2345}}{(k_{35})(k_{34.5})(k_{23.45})(k_{13.245})}X_3 \\ & + r_{14.235}\frac{\sigma_1}{\sigma_4}\frac{k_{1.2345}}{(k_{24})(k_{34.2})(k_{45.23})(k_{14.235})}X_4 \\ & + r_{15.234}\frac{\sigma_1}{\sigma_5}\frac{k_{1.2345}}{(k_{25})(k_{35.2})(k_{45.23})(k_{15.234})}X_5 - C \qquad \text{(B)}\end{aligned}$$

If, in terms of the illustrative problem, we let X_1 = weight, X_2 = height, X_3 = chest depth, X_4 = chest breadth and X_5 = hip width, then it can readily be seen that all the partial r's and partial k's in the first two terms of equation (B) are taken care of by the computations on Diagram IV and Diagram V. From Diagram IV we get k_{15}, k_{25}, k_{35}, and from Diagram V we get $k_{24.5}$, $k_{34.5}$, $k_{23.45}$, $r_{12.345}$ and $r_{13.245}$.

In order to compute the partials necessary for the two remaining terms it is necessary to make the same number of computations as those made in Diagrams IV and V, reversing the order in which the variables are partialled out; in terms of the illustrative problem, partial out the variables in the order height, chest depth and chest breadth in addition to the order hip width, chest breadth and chest depth as illustrated on the diagrams presented. Such computations would yield the partials required for the remaining two terms of equation (B).

In problems involving more than 5 variables it is only necessary to go from zero order r's to the $(n - 2)$ order partials one time in order to obtain the Multiple R. If the regression equation is desired it is necessary to compute four $(n - 2)$ order partials in the manner indicated above and then return to the second order r's, rearrange them on the work sheet and then compute other partials of order $(n - 2)$. For example, if the problem involves 7 variables, and in the first computation the variables are partialled out in the order 7, 6, 5, 4, 3, giving the two partials $r_{12.34567}$ and $r_{13.24567}$, and if in the second computation the variables are partialled out in the order 2, 3, 4, 5, 6, giving the partials $r_{17.23456}$ and $r_{16.23457}$, the remaining two partials desired could be obtained by partialling out in the order 2, 3, 7, 6, 4, giving $r_{14.23567}$ and $r_{15.23467}$. But the computations necessary to partialling 2 and 3 out were done in the second computation, so a rearrangement of the sequence on the computational form of the second order r's is all that is necessary for these additional higher order r's. This again reduces the amount of computation necessary.

Though the explanation may seem long and complex, the routine is very simple. After it is understood and has become mechanical in operation, the time saving is approximately one-half of the usual method. It can be explained to clerks verbally much more easily than has been possible in this written explanation.

A STUDY OF QUESTIONNAIRE TECHNIQUE

FRANK K. SHUTTLEWORTH

Yale University

This study shows that under the conditions defined below the enclosure of a twenty-five cent piece in a mail questionnaire brought 32.4 per cent more replies than the same questionnaire without the coin. Six hundred eight persons receiving the coin returned 51.6 per cent replies while three hundred seventy-six not receiving the coin returned only 19.1 per cent. Data are also presented on the rôle of selective factors and on relative costs.

The study is a by-product of an evaluation of the Cattaraugus County Health Demonstration in western New York. For a period of seven years the Milbank Memorial Fund financed an intensive demonstration of rural public health work in Cattaraugus County. In the spring of 1930 a staff of experts under the direction of Professor C-E. A. Winslow[1] undertook to evaluate that work. The author assisted in a survey of the educational aspects of the demonstration among school children and adults. With some reluctance a questionnaire by mail was used to assay the attitudes of adults especially toward the financial support of public health work. One of our chief doubts about the questionnaire was the fear of an inadequate proportion of replies. To solve this difficulty, it was decided to enclose a twenty-five cent piece secured in a coin mailing card with each questionnaire. Evaluation of the results required a control county and for this purpose, Steuben, the county just east but one from Cattaraugus, was selected. The inquiry to these counties consisted of a short personal letter each of which was separately typed and obviously so, the questionnaire proper containing six questions, a stamped and addressed envelope, and the coin.

The investment involved and the novelty of the procedure naturally led to an attempt to test the effect of the coin. Accordingly, the inquiry was duplicated in all respects save only two sentences in the personal letter and the presence of the coin. This inquiry was sent to persons in the six New York counties immediately surrounding Cattaraugus and Steuben.

The personal letter which accompanied the coin was as follows:

---

[1] Winslow, C-E. A., "Health on the Farm and in the Village" Macmillan, 1931.

A responsible organization has asked me as an impartial outsider to make a survey of opinion toward public health work in the western counties of New York State. They are anxious to have 100 per cent replies to the questions and accordingly have provided funds which make possible the enclosed coin. It is not in payment for your trouble in answering the questions but rather a small token of appreciation of your cooperation.

The letter not accompanied by a coin substituted the following for the last two sentences in the above letter:

The results of such a survey will be trusted to the extent that every one cooperates in answering the enclosed questions. I will be grateful if you will indicate your opinion and send it to me quite promptly in the enclosed stamped and addressed envelope.

These letters were written on the stationery of the Department of Education of Yale University.

The comparability of the two areas to which these inquiries were sent is of importance in evaluating the influence of the coin. In connection with the health study, which was searching for differences between Cattaraugus and Steuben, all available population, economic, and educational statistics were examined. On thirty-seven basic factors the two counties proved to be very much alike. Cattaraugus is, of course, very different from Steuben in one respect having had a well organized and adequately financed county health program for seven years while Steuben had had no organized county health program. However, the replies to the questionnaire showed only slender advantages on the part of the Cattaraugus respondents. For the purpose of this study we feel justified in combining them into one area which will be referred to as the coin area. Additional data testing the comparability of the coin area and the non-coin area was not examined for the reason that only the portions of the six counties most immediately adjacent to Cattaraugus and Steuben were canvassed. Instead, reliance is placed on the similarity between Cattaraugus and Steuben which are separated by the intervening county of Allegheny. The complete details are recorded in the accompanying table in the order in which they are discussed in the text. Data for the coin area are also given separately for Cattaraugus and Steuben counties.

The inquiry itself provided other controls. The questionnaires were sent only to names listed in village telephone directories, thus introducing an economic control, and approximations to social and cultural controls. In the non-coin area sixteen village exchanges

serving forty-two still smaller hamlets and cross-roads were sampled. In the coin area thirty village exchanges serving fifty-seven additional hamlets were sampled. The sixteen villages in the non-coin area average 8.6 miles from the borders of the coin area. All of the questionnaires were mailed on the same day.

Questionnaires were mailed to every eighth name on these exchanges excepting the names of business firms and of women. This gave three hundred eighty letters to the non-coin and six hundred seventeen to the coin area. Four and nine letters were returned for insufficient address. Of the three hundred seventy-six and six hundred eight persons who presumably received the inquiry, seventy-two or 19.1 per cent from the non-coin area and three hundred fourteen or 51.6 per cent from the coin area made some reply. The difference of 32.4 per cent is sixteen times as large as its probable error.

The coin area returned 2.7 times as many replies. This ratio gradually increased throughout the period during which replies were coming in. The first three days with a total of one hundred fifty-three replies showed a ratio of 2.5, the second three days with one hundred fifteen replies showed a ratio of 2.6, while the second, third, and fourth to tenth weeks showed ratios of 2.7, 3.9, and 4.9.[1]

A detailed analysis of the responses is presented since it throws light on two problems of the questionnaire method. One of the haunting ghosts of this procedure is the selection which presumably is involved in the fact that some persons reply and others do not. Presumably less selection is involved in 52 per cent replies than in 19 per cent. On the other hand there is the question whether the presence of the coin tended to disturb the answers. The data offer many tests of these questions although in the nature of the case it is impossible to say whether one or both factors are responsible for any differences. Interpretation of similarities is equally difficult since they may mean that neither selection nor the coin disturbed the answers or that these two factors tended to cancel each other.

A preliminary point of some interest is that fourteen of the three hundred fourteen respondents from the coin area returned the coin. Seven of these filled out the questionnaire in detail and by comments or implication indicated that they thought the coin unnecessary while

[1] Subsequent to the preparation of this article ten weeks after mailing the questionnaires an additional ten replies were received up to the end of the fourth month. Nine of those were from the coin area and one from the non-coin area. These additional cases are not included in this analysis.

seven returned the coin without the questionnaire or with the questionnaire left blank. Although the questionnaire said specifically "You do not need to sign your name" twelve of the fourteen or 86 per cent gave their names while among the other three hundred respondents only fifty-two or 17 per cent gave their names.

The sixth question in the inquiry asked whether the respondent's taxes on real estate amounted to over $100 the previous year. In the non-coin area, of sixty-eight answering this question, 63.2 per cent reported more than $100 taxes. In the coin area of two hundred eighty-five answering this question 59.5 per cent made this report. The difference is 3.7 $\pm$ 4.6 per cent. All of the data were tabulated separately for those reporting more and for those reporting less than $100 taxes but since no differences in the answers of the two groups appeared they are combined in the discussion which follows.

The first question in the inquiry asked the respondents to rank four functions of public health work in order of importance. Forty-eight from the non-coin area and one hundred fifty-eight from the coin area properly followed instructions in ranking the four functions. The percentages of correct rankings for function (*a*) were 47.9 and 44.3 per cent, for function (*b*) 81.3 and 77.9 per cent, for function (*c*) 27.1 and 36.7 per cent, for function (*d*) 54.2 and 51.9 per cent. The largest difference here is 9.6 $\pm$ 5.3 per cent.

The second, third, fourth, and fifth questions concerned financial support of public health work. Favorable attitudes were expressed by answering "No" to the second question and "Yes" to the third, fourth, and fifth. For the non-coin and coin areas the percentages of favorable replies were 79.4 and 77.4 per cent, 70.5 and 70.4 per cent, 41.9 and 39.5 per cent, and 73.8 and 77.2 per cent. Here the largest difference is only 3.4 per cent.

We turn next to the care with which the inquiry was answered by the two areas as indicated by the proportion of omitted or uncertain answers. Instead of seventy-two and three hundred fourteen cases as our base we shall use seventy-one and three hundred seven since one respondent from the non-coin area wrote that he had mislaid his questionnaire and seven from the coin area returned the coin without answering any questions. For the six questions the non-coin area averages 15.0 per cent omissions or uncertain answers while the coin area averages 17.0 per cent. Analysis by specific questions, however, reveals one significant difference: Only 32.9 per cent from the non-coin area failed to rank the four functions of public health as instructed in

comparison with 50.4 per cent from the coin area. The difference of 17.5 per cent is 4.1 times as large as its probable error. On two other questions the coin area showed more omissions while on the other three it showed fewer omissions. The largest of these differences is only 6.2 per cent.

The respondents from the two areas were equally ready to back their replies with their signatures. Although the questionnaire specifically said, "You do not need to sign your name," 15.5 per cent from the non-coin area signed their names. After deducting the fourteen cases who returned the coin and naturally wanted to receive credit for it there remain 18.0 per cent of the respondents from the coin area who gave their names.

A final test of the influence of selection and of the coin is the interest indicated by comments. These were invited by "Use the reverse side for comments" following the sixth question. From the non-coin area 44.5 per cent made comments while from the coin area only 33.2 per cent made any comments. The difference of 11.3 per cent is only 2.6 times as large as its probable error. While a smaller proportion of respondents from the coin area made comments, those who did so made more and longer comments. Counting each qualification to a specific question and each paragraph of discussion as a comment, the non-coin area averages 1.78 comments per person commenting and .79 comments per person replying. In the coin area these figures are 1.92 and .64. Counting all words in the comments gives an average of 30.40 per person commenting and of 13.50 per person replying in the non-coin area in comparison with 48.92 and 16.25 in the coin area. The most significant of these differences is only 2.2 times as large as its probable error. The differences are also exaggerated because these measures are the ones in which Cattaraugus approaches significant superiority over Steuben.

The relative costs of the two procedures are of some interest. No expense was spared to give the inquiry a favorable appearance. The letters to the non-coin area cost 10.4 cents each while those to the coin area cost 36.5 cents. The initial expense to the coin area was 250 per cent greater per letter. The larger proportion of replies reduced this discrepancy very much. In costs per reply of any kind, the coin area exceeded the non-coin area by only 31 per cent. In costs per relatively complete reply the excess for the coin area was only 32 per cent. In costs per reply containing one or more comments the excess was 80 per cent. In costs per comment the excess was 61 per cent.

In costs per word of comment the coin area exceeded the non-coin area by only 12 per cent.

Summary

The enclosure of a twenty-five cent piece in a simple questionnaire sent out by mail brought 51.6 per cent replies while the same questionnaire without the coin returned only 19.1 per cent replies.

Table I.—Summary of Data

	Non-coin		Coin		Cattaragus		Steuben	
	No.	Per cent or *M*	No.	Per cent or *M*	No.	Per cent or *M*	No.	Per cent or *M*
Village telephone exchanges sampled	16	.....	30	.....	14	.....	16	
Other hamlets included	42	.....	57	.....	23	.....	42	
Inquiries mailed	380	.....	617	.....	287	.....	330	
Returned for address	4	.....	9	.....	2	.....	7	
Persons receiving inquiry	376	.....	608	.....	285	.....	323	
Replies of any kind	72	19.1	314	51.6	151	53.0	163	50.5
Wrote that questionnaire was mislaid	1							
Returned coin with questions answered	...	.....	7	.....	6	.....	1	
Returned coin without answering questions	...	.....	7	.....	4	.....	3	
Useable returns	71	18.7	307	50.5	147	51.6	160	49.5
Reported less than $100 taxes	25	36.8	124	43.5	54	38.3	70	47.9
Reported more than $100 taxes	43	63.2	161	59.5	85	61.2	76	52.1
Correctly ranking function (*a*) fourth	23	47.9	70	44.3	39	47.0	31	41.3
Correctly ranking function (*b*) first or second	39	81.3	123	77.9	64	77.1	59	78.7
Correctly ranking function (*c*) third	13	27.1	58	36.7	29	34.9	29	38.7
Correctly ranking function (*d*) first or second	26	54.2	82	51.9	45	54.2	37	49.3
Correctly answering question No. 2 "No"	50	79.4	215	77.4	98	76.5	117	78.1
Correctly answering question No. 3 "Yes"	43	70.5	197	70.4	97	74.6	100	66.7
Correctly answering question No. 4 "Yes"	26	41.9	111	39.5	55	41.7	56	37.8
Correctly answering question No. 5 "Yes"	45	73.8	220	77.2	111	81.1	109	71.8
Uncertain or omitted answers to six questions	63	15.0	313	17.0	156	17.7	157	16.3
Uncertain or omitted tax report	3	4.3	29	9.5	12	8.2	17	10.6
Uncertain or omitted ranking of functions	23	32.9	156	50.4	68	46.3	88	55.0
Uncertain or omitted answers to question No. 2	8	11.4	36	11.7	23	15.6	13	8.1
Uncertain or omitted answers to question No. 3	10	14.3	34	11.1	21	14.3	13	8.1
Uncertain or omitted answers to question No. 4	9	12.8	33	10.7	18	12.2	15	9.4
Uncertain or omitted answers to question No. 5	10	14.3	25	8.1	14	9.5	11	6.8
Persons signing names	11	15.5	66	.....	36	.....	30	
Deducting fourteen cases returning coin, twelve of whom signed names	...	.....	54	18.0	28	17.7	26	16.2
Persons making comments	32	44.5	102	33.2	53	36.1	49	30.6
Average number of comments per person commenting	...	1.78	...	1.92	...	2.21	...	1.61
Average number of words per person commenting	...	30.40	...	48.92	...	57.34	...	39.82
Average number of comments per person replying	...	.79	...	.64	...	.80	...	.49
Average number of words in comments per person replying	...	13.52	...	16.25	...	20.67	...	12.19
Total expense of sending inquiry in dollars	...	39.40	...	225.50	...	104.89	...	120.61
Costs per letter sent out in cents	...	10.4	...	36.5	...	36.5	...	36.5
Costs per reply of any kind in cents	...	54.7	...	71.8	...	69.5	...	74.0
Costs per relatively complete reply in cents	...	55.5	...	73.5	...	71.4	...	75.4
Costs per reply with any comment in cents	...	123.0	...	221.1	...	178.9	...	246.1
Costs per comment in cents	...	69.1	...	115.1	...	89.7	...	252.7
Costs per word of comment in cents	...	4.0	...	4.5	...	3.5	...	6.2

Detailed analysis of the returned questionnaires shows negligible difference between the non-coin and coin areas. On one question the respondents from the coin area showed significantly more omissions or uncertain answers, but considering omissions and uncertain answers to all questions the advantage of the non-coin area is only 2 per cent. A larger proportion of respondents from the non-coin area made comments, but the comments of respondents from the coin area tended to be longer.

From the absence of differences in the responses of the non-coin areas to the questionnaire, one of the following inferences is warranted: Neither the selection involved in a small proportion of replies nor the presence of the coin was disturbing to the replies. The influence of selection and of the coin cancelled each other.

While the initial cost of enclosing a twenty-five cent piece made the letters to the coin area 250 per cent more expensive than letters to the non-coin area, the larger returns from the coin area reduced the excess in terms of unit costs as low as 12 per cent. At the most the actual excess expense is 80 per cent per unit of returns.

The procedure of paying the victims of a questionnaire is recommended, not for adoption, but for serious consideration. In terms of costs it did not prove superior in this instance. Whether on the whole it was superior in the present case is a matter of judgment which must go beyond the actual data in evaluating the importance of selective factors and the possibility that the coin itself was a disturbing factor.

INFLUENCE OF THE ASSIGNMENT ON LEARNING

DAVID H. BRIGGS

Florida State College for Women

A. M. JORDAN

University of North Carolina

The more effective utilization of the assignment as an instrument of instruction has been the burden of much discussion but of little experimentation. Recent writers such as Odell, Crawford and McDonald, Stone, Sears, and Monroe have pointed out the need for adequate assignments of lessons. By "adequate" they imply that assignments should be clear, helpful as to study procedures, stimulating, and that they should arouse problems to be solved. With the outcomes of these procedures they have not concerned themselves. It has seemed to us worthwhile, therefore, to investigate the amount of added learning which accrues from certain assignment procedures.

The assignment procedures, five in number, were chosen because of their general use rather than because they are the best. These procedures are:

(*a*) The pupils were furnished a suggested study procedure which consisted of looking up meaning of difficult words, drawing lines under the important statements, and learning the significance of the statements previously underlined. Great care was taken to make sure that the pupils knew exactly what they were to do before they were allowed to begin.

(*b*) Carefully selected questions on the material to be studied were given to the pupils before they began studying. The instructions were "I want you to find the answers to these questions, and when you have found them draw a line under the correct answer. Try to learn as many other facts as you can while you are reading and studying the selection. You are to use these questions as a guide in helping you to master the points brought out in the selection."

(*c*) The pupils were taught the meaning of the difficult words in the material to be studied. A sheet was presented to the pupils containing a list of the more difficult words with their appropriate meanings with these instructions: "After we have spent about ten minutes studying these hard words I am going to give you a test to see how many of them you have learned. Then, too, you will need to know their meanings in order to answer the questions you will be asked later on in the passage you are to read."

(*d*) Attempt was made to show the children the manner in which the tariff (the subject of the material studied) affected them personally. Illustrations were presented of types of goods made more cheaply in foreign lands than we could manufacture in this country because of the lower standards of living obtaining

there than here. The ordinary high tariff argument was presented which emphasized the effect of lowering or raising the tariff on wages and family income and the consequent effect on the pupils themselves.

(*e*) In this case, there was a conscious attempt to explain what the tariff was, why some thought it a good measure, why it is called a protective tariff and why it was levied. It was emphasized that many of the pupils might differ from the President politically but that it was their duty to learn much about a topic discussed in Congress at that very time.

The material used was selected from the message of President Hoover on the general subject of the tariff. It consisted of two pages of single spaced typewritten material and was in nature more difficult than the ordinary reading matter used by pupils of the levels studied. The four pages were divided equally into two sections, *A* and *B*. This could be done rather easily and naturally since there was a break in the thought at the point of division. For each paragraph there were prepared three sorts of material. The first consisted of a series of words chosen from the more difficult words of the text with multiple choice answers. The second type of material was a test of sentence meaning. Care was exercised in making the answers depend on the understanding of the individual sentence and of that alone. The third type consisted of statements based on the interrelation and meaning of several sentences taken together. Sometimes these questions depended for their answer upon the interpretation of what was read. These three types are called: (1) Word knowledge, (2) sentence meaning, (3) paragraph meaning. Before the experiment proper, this material was checked for its difficulty (1) by counting the number of words in each selection more difficult than the first five thousand in "Thorndike's Word Book" and (2) by testing the material on two hundred forty-one cases located in Grades V, VII, and IX. This last procedure furnished us with the following results:

	Word knowledge	Sentence meaning	Paragraph meaning
Form A	13.11	6.06	4.47
Form B	14.03	5.54	4.16
Differences of means	.92	.52	.32
$PE_{diff.}$ of means	.37	.14	.12

The PE of the differences of the means showed the differences between *A* and *B* in sentence meaning and in paragraph meaning to be reliable. The difference was unreliable in the case of word knowledge.

The differences between the means were provided for by presenting *A* and *B* in alternate manner. If there were two grades, in the one we used *B* for the control group and *A* for the experimental group; in the other, *A* was the control and *B*, the experimental. The correlation between the two forms was .86 ± .01.

General Procedure

The five situations were tested in the Durham and Raleigh, N. C., schools. Two groups of fifth, two of seventh, and two of ninth grade pupils were subjects for each of the experimental situations. This made a total of one hundred sixty-seven subjects for Assignment *A*, one hundred eighty-six for Assignment *B*, two hundred fifteen for Assignment *C*, one hundred eighty-nine for Assignment *D*, and one hundred ninety-four for Assignment *E*—nine hundred fifty-one subjects in all. The same examiner (Briggs) gave all the assignments and all the tests to each of the thirty groups of pupils.

The general plan was to test the children first with one of the forms on one day, then the next day make the assignment with the other form and calculate the differences scored by pupils who on one day were influenced by certain types of assignments and on the other were not so influenced. The total time consumed by the pupils in these procedures was from thirty-five to fifty minutes.

Results

The average results only are given on account of the lack of space.

Table I.—Results for Assignment A
(Suggested Study Procedure)

	Word knowledge	Sentence meaning	Paragraph meaning
Control	15.58	6.34	4.73
Experimental	18.66	6.43	4.80
Number	167.	167.	167.
Difference of means	3.08	.09	.07
$PE_{diff.}$ of means	.53	.23	.19

In studying Table I, one is struck by the slight differences between the means. In the sentence meaning and in paragraph meaning the differences are merely chance differences but in word knowledge there was a reliable gain due to the assignment. It is to be remembered

that the very procedure of Assignment *A* almost forced an improvement in vocabulary since it consisted among other things of looking up the meaning of difficult words. Since we had at our disposal the records of several grades it was not difficult to get an average gain for the year in each of the tests. The average gain per year on the vocabulary test was 3.07 words which is just about the same as the improvement of the vocabulary scores as a result of this procedure. The mean gain per year for sentence meaning was .81 questions and for paragraph

TABLE II.—RESULTS FOR ASSIGNMENT *B*
(Assignment of Pertinent Study Questions)

	Word knowledge	Sentence meaning	Paragraph meaning
Control	16.48	6.33	4.78
Experimental	16.33	6.53	4.88
Number	186	186	186
Difference of means	−.15	.20	.10
$PE_{diff.}$ of means	.47	.21	.17

meaning .75 questions. Although the meanings of the words were studied and learned sufficiently well to recognize their correct meaning among five words they were not learned sufficiently well in the time allowed for them to have entered into the warp and woof of the sentence in such a manner as to improve the scores on the sentences and paragraphs.

TABLE III.—RESULTS OF ASSIGNMENT *C*
(Teaching Meaning of Difficult Words)

	Word knowledge	Sentence meaning	Paragraph meaning
Control	15.06	6.34	4.74
Experimental	26.69	6.60	5.03
Number	215	215	215
Difference of means	11.63	.26	.29
$PE_{diff.}$ of means	.41	.20	.16

It is clear from Table II that the assignment of pertinent study questions had little if any demonstrable result on word knowledge or paragraph meaning and a slight but unreliable effect on understanding the meaning of sentences.

It is apparent that the words learned in the experimental group did influence the scores in word knowledge. More interesting perhaps is the small improvement in sentence and paragraph meaning when so many of the words had been learned. The improvement looms larger, however, when compared with the yearly growth of .81 questions in the sentence meaning and of .75 in paragraph meaning. The difference between the experimental and the control in word knowledge needs no comment since these words were taught the children directly. In

TABLE IV.—RESULTS OF ASSIGNMENT *D*
(Making Pupils Aware of Personal Value of Materials)

	Word knowledge	Sentence meaning	Paragraph meaning
Control	17.87	6.96	5.43
Experimental	18.57	6.92	5.48
Number	189	189	189
Difference of means	.70	−.04	.05
$PE_{diff.}$ of means	.49	.22	.18

sentence meaning the chances are sixty-two in a hundred that if the experiment were repeated the experimental group would still be ahead of the control group. In the case of paragraph meaning similar chances are seventy-eight in a hundred. Teaching the meaning of difficult words, as far as our study goes, has a decided effect upon the ability to comprehend the meaning of sentences and paragraphs.

TABLE V.—RESULTS OF ASSIGNMENT *E*
(Making Pupils Aware of the General Background of the Material)

	Word knowledge	Sentence meaning	Paragraph meaning
Control	17.32	6.77	5.48
Experimental	18.09	7.07	5.59
Number	194	194	194
Difference of means	.77	.30	.11
$PE_{diff.}$ of means	.47	.20	.16

As a result of making pupils aware of the personal value of the tariff, there was a fairly significant improvement in word knowledge. The effects on sentence meaning and paragraph meaning were nil.

Making children aware of the general background of the material studied produces a positive effect in the three types of answer required of them. The differences are not statistically reliable but all of them are positive and in the same direction. Compare these differences with a year's growth of 3.08 in word knowledge, .81 in sentence meaning and .75 in paragraph meaning and it is clear that a few minutes of explanation brings about ¼ of a year's growth in word knowledge, ⅓ of a year's growth in sentence meaning, and ⅐ of a year's growth in paragraph meaning. It will be seen that the improvement is greatest in sentence meaning and least in paragraph meaning.

TABLE VI.—MEAN GAINS IN EACH OF THE SITUATIONS WITH THEIR PROBABLE ERRORS

	Word knowledge	Sentence meaning	Paragraph meaning	Number
(*a*) Suggested study procedure	3.08 ± .53	.09 ± .23	.07 ± .19	167
(*b*) Pertinent study questions.	−.15 ± .47	.20 ± .21	.10 ± .17	186
(*c*) Teaching meaning of words	11.63 ± .41	.26 ± .20	.29 ± .16	215
(*d*) Making pupils aware of personal value of materials	.70 ± .49	−.04 ± .22	.05 ± .18	189
(*e*) Developing background...	.77 ± .47	.30 ± .20	.11 ± .16	194
Average yearly gain.....	3.08	.81	.77	

Table VI summarizes the results of this investigation. It will be seen that the scores encircled are as large or larger than their corresponding probable errors. From the strictest statistical point of view only Assignments *A* and *C* are reliable for in these cases the differences are more than four times the probable error of the difference. But in these cases the teaching directly influenced the scores on Word knowledge. It is clear that the teaching of the meaning of words produced an improvement in sentence meaning of .26 of a question or a little less than ⅓ of a year's growth and slightly more than ⅓ of a year's growth in paragraph meaning. This careful teaching of the meaning of the more difficult words stands out as one good way to make an assignment. Another method almost as effective is the development of a background for understanding the material to be studied. This background of understanding is of undoubted assistance in comprehending the meaning of the material studied.

Since these assignments were tried out upon Grades V, VII, and IX, opportunity was offered to study the differential effect of the assignment upon each grade.

TABLE VII.—COMPARATIVE EFFECTIVENESS OF THE FIVE METHODS OF MAKING THE ASSIGNMENT ON FIFTH, SEVENTH, AND NINTH GRADE PUPILS

	Grade V	Grade VII	Grade IX
Word knowledge	C	C	C
Sentence meaning	E	E	B
Paragraph meaning	D	E	B

Teaching the meaning of the difficult word (Situation C) caused the greatest gain in word knowledge for each of the three grades tested. Making pupils aware of the general meaning of the materials (Situation E) caused the greatest gain in sentence meaning on the part of fifth and seventh grade pupils while with the ninth grade it was the assignment of pertinent questions (Situation B). In developing paragraph meaning, making the pupils aware of the personal value of materials appears to be best in the fifth grade; while making pupils aware of the general background of the material studied is most successful in the seventh grade.

While the gains in the experimental situations over the control situations are not statistically significant, except in two cases of word knowledge (Situations A and C), yet our data show strong indications of the influence of the assignment. The consistency of the gain, the fact that each type caused a noticeable gain on the part of one or two of the grades examined, the brief time for acquainting the pupils with the specific procedure used in the assignment, and the large per cent of a year's gain in many cases, are evidences of the importance of the assignment.

REFERENCES

Book, W. F.: The Development of Higher-Orders of Perceptional Habits in Reading. *Jour. Educational Research*, Vol. XXI, No. 3, March, 1930.

Crawford, C. C. and McDonald, Lois P.: "Modern Methods of Teaching Geography." Boston: Houghton Mifflin Company, 1929, pp. 238–240.

Jordan, A. M.: *Educational Psychology*. Henry Holt & Co., 1928, Chap. IV.

Monroe, W. S.: "Directing Learning in the High School." New York: Doubleday Doran & Co., 1928, pp. 418–420.

Monroe, W. S. and Mohlman, Dora K.: Training in the Technique of Study. *University of Illinois Bulletin*, Vol. XXII, No. 2, *Bureau of Educational Research Bulletin* No. 20, Urbana, 1924.

Odell, Charles: The Assignment of Lessons. *University of Illinois Bulletin*, Vol. XXIII, No. 7, Oct. 19, 1925, p. 4.

Sears, J. B.: "Classroom Organization and Control." Boston: Houghton Mifflin Co., 1929, p. 229.

Stone, C. R.: "Supervision of the Elementary School." Boston: Houghton Mifflin Co., 1929, p. 341.

Waples, D.: "Procedure in High School Teaching." Macmillan Co., 1924.

FACTUAL MEMORY OF SECONDARY SCHOOL PUPILS FOR A SHORT ARTICLE WHICH THEY READ A SINGLE TIME

ALFRED G. DIETZE AND GEORGE E. JONES

University of Pittsburgh

(*Concluded from November Issue*)

FORGETTING

Comparison of Group Averages.—Table VII repeats the averages presented in Tables I to V[1] in such a manner as to reveal the influence of lapse of time to bring about forgetting of an article read a single time. It is clear from these figures that with passage of time pupils remember fewer and fewer of the facts contained in the articles. Grade by grade and article by article eighty-four comparisons may be made between successive intervals. With six exceptions, only one of which is large enough to be significant, the averages of any given interval are smaller than those of the intervals preceding. This is noteworthy in view of the fact that the figures are based on groups of varying abilities.

Table VIII gives the A-score averages for the combined groups.[2] Average scores decrease consistently with time in the figures of this table. The conclusion seems warranted that, within the limits of the present study, the greater the interval between the reading of an article and the subsequent attempt to recall its content the greater is the forgetting for the facts contained.

Results of Parallel Groups.—The above findings were checked for accuracy by using the method of equivalent groups. One hundred cases were selected from each of the six experimental groups on the basis of A-scores in immediate memory in such a manner that the six resulting groups were exactly equivalent and that the A-scores of each were normally distributed. Since 56.7 happened to be the A-score average of the combined grades, we selected 56.0 as the mean and 8.6 as the standard deviation of the new parallel groups and then computed the theoretical frequencies for six class intervals of four

[1] See first half of this article in previous issue.

[2] It should be recalled that the A-score average represents the work of all groups taken together in terms of the material Arkwright.

points each above and below the mean in a normal curve with mean and standard deviation as above and area equal to one hundred. The

TABLE VII.—AVERAGE MEMORY AFTER DIFFERENT INTERVALS[1]

Recall interval	Grades						
	VII	VIII	IX	X	XI	XII	All
Germans							
Immediate	53.8	57.9	67.3	66.6	73.0	75.8	64.8
One day	49.0	55.1	54.2	58.8	59.3	71.8	56.7
Fourteen days	37.9	34.3	38.8	44.2	46.7	51.3	42.0
Thirty days	32.9	*41.3*	32.1	32.1	40.1	39.3	36.4
One hundred days	....	29.4	26.5	31.4	35.9	33.3	32.4
Radium							
Immediate	67.3	66.6	72.4	76.6	77.0	84.3	73.4
One day	47.7	55.0	56.1	56.3	64.9	72.9	58.7
Fourteen days	40.7	45.2	44.9	40.8	50.6	40.1	43.7
Thirty days	37.2	37.7	37.1	*42.2*	42.1	37.9	38.3
One hundred days	27.3	31.9	*38.6*	36.6	35.1	*39.1*	33.8
Arkwright							
Immediate	43.9	50.0	50.3	57.4	62.7	64.8	54.6
One day	33.5	43.4	42.4	49.7	55.2	63.3	46.4
Fourteen days	*37.5*	28.8	32.9	34.4	32.3	32.0	33.1
Thirty days	25.9	27.6	25.4	*37.8*	31.5	31.3	30.3
One hundred days	20.5	22.0	21.9	28.7	23.0	22.6	23.4

[1] Italicized figures show scores that are higher than preceding interval.

TABLE VIII.—AVERAGES MEMORY AFTER DIFFERENT INTERVALS—A-SCORE

Recall interval	Grades						
	VII	VIII	IX	X	XI	XII	All
Immediate	50.0	51.3	55.5	59.0	62.3	65.2	56.7
One day	40.0	45.3	46.0	47.9	53.2	61.1	48.0
Fourteen days	36.5	33.8	35.9	36.1	39.0	36.8	36.3
Thirty days	25.5	33.2	29.5	35.7	34.9	33.8	32.4
One hundred days	25.5	27.9	28.5	30.5	29.6	31.2	30.0

hundred cases of each group were then taken to fit these frequencies. An attempt was made to keep the number from each grade approximately the same.

The reader should note that this method makes possible the comparison of different intervals on the basis of a given material even though a different group was used for each interval;—*e.g.*, in the case of Radium we may compare the results of Group I or II in immediate memory with Group V in memory after one day, Group VI in memory after fourteen days, Group IV in memory after thirty days, and Group III in memory after one hundred days. The six groups are almost exactly alike in terms of A-scores in immediate memory.

The results for each material are presented in Table IX. The reader should note especially the close correspondence between the A-score column of the present table with the last column of Table VIII. Such close correspondence between two methods of analysis is an indication of the validity of the methods.

TABLE IX.—AVERAGE MEMORY OF PARALLEL GROUPS AFTER DIFFERENT INTERVALS[1]

Interval	Radium	Germans	Arkwright	A-score
Immediate	66.0	67.7	56.0	56.0
One day	61.3	49.4	47.8	47.8
Fourteen days	45.0	37.7	33.0	35.6
Thirty days	41.2	36.9	26.2	32.4
One hundred days	35.4	32.7	22.4	28.7

[1] Each average is based on one hundred cases, the A-scores on three hundred.

Rate of Forgetting.—A curve of forgetting may be plotted either from the data of the last column in Table VIII or from those of the last column in Table IX. Such a curve would be representative of the forgetting of pupils in Grades VII to XII for the article "Arkwright" after a single reading. The data of Table VIII are plotted in the forgetting curve of Fig. 3. The curve is begun at sixty-three per cent, since this is approximately the amount of the test which pupils in these grades can answer when they are permitted to look up the answers in the text,—*i.e.*, sixty-three per cent of the article "gets across" in a single reading (see *supra.* paragraph entitled, "Comprehension Difficulty and Carelessness"). To enable the reader also to determine how much of what actually "gets across" in a single reading is remembered at each of the intervals, a second scale is provided to the right of the graph which begins the curve at one hundred per cent. The asymptote drawn in represents the estimated point of complete forgetting for the material, which is taken at six per cent since it was

found that pupils could answer this amount of the test without previously having read the article (see *supra.*, paragraph entitled "Previous Knowledge").

Summarizing and interpreting the data presented, we find that pupils in the grades studied are able to answer an average of sixty-three per cent of the questions on Arkwright when they are permitted to look up the answers in the article; that immediately after the reading they remember an average of 56.7 per cent; after one day an average of 48.0 per cent; after fourteen days an average of 36.3 per cent; after

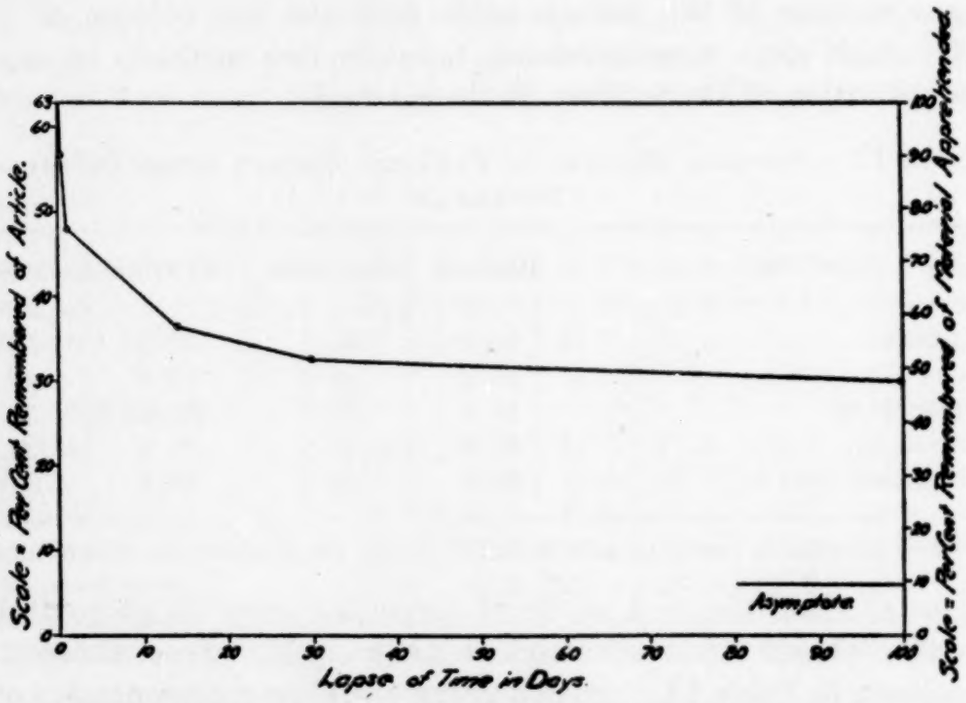


FIG. 3.—Average forgetting curve for Arkwright based on A-Scores of all groups participating in the experiments.

thirty days an average of 32.4 per cent; and after one hundred days an average of 30.0 per cent. The immediate memory score represents an average loss of 43.3 per cent of the facts presented in the article read; of this, the difference between one hundred per cent and sixty-three, or thirty-seven per cent, is due to failure to comprehend and to careless working habits, and the remainder, 6.3 per cent, to forgetting during the interval of reading the article and taking the test. If we compute the per cent remembered of that which actually "gets across," we obtain scores as follows: Immediate memory 90.0 per cent, memory after one day 76.2 per cent, memory after fourteen days 59.3 per cent,

memory after thirty days 51.7 per cent, and memory after one hundred days 47.6 per cent.

We find here the same principle discovered by Ebbinghaus,—*viz.*, that in forgetting there is a rapid initial loss and a more gradual subsequent loss. We find also the principle corroborated that meaningful material is forgotten less quickly than nonsense syllables; and we may add that facts learned in a single reading of an article of the difficulty of Arkwright are, in the longer intervals at least, forgotten less rapidly than in meaningful material in the form of poetry learned to the degree of one correct and unhesitating recital. In Table X we present data taken from three classic studies of forgetting[1] and compare them with our results. A study of the table shows that the amounts forgotten of what is actually apprehended of Arkwright are less, in general, than for the other materials, especially after intervals of more than one day. From the position of the asymptote in Fig. 3, it is also apparent that the point of complete forgetting is still far off even after one hundred days, and, from the shape of the curve, that it would take a long time to reach this point, if it were ever reached.

TABLE X.—MEMORY AFTER DIFFERENT INTERVALS
Comparison of Results of Ebbinghaus, Radossawljewitsch and Finkenbinder with Present Study

Interval	Nonsense syllables			Poetry	Factual
	Ebbinghaus	Radossawljewitsch	Finkenbinder	Radossawljewitsch	Dietze
Immediate (20 minutes)	58	89	23[2]	96	90
One day	34	68	58	79	76
Fourteen days	22[1]	41	47[2]	30	59
Thirty days	21	20	44[2]	24	52
One hundred days	?	4[1]	38[2]	?	48

[1] By interpolation in author's tables.
[2] By entry in Finkenbinders tables of 10 Log.

[1] Ebbinghaus, H.: *Ueber das Gedaechtnis.* Leipzig, 1885: translated by H. A. Ruger and E. Bussenius, *Memory.* Bureau of Publications, Teachers College, Columbia University, New York, 1913.

Radossawljewitsch, P. R.: *Das Fortschreiten des Vergessens mit der Zeit.* Druck der Dieterich'schen Universitaet, Buchdruckerei, Goettingen, 1905.

Finkenbinder, E. O.: The Curve of Forgetting. *American Journal of Psychology,* Vol. XXIV, 1913, pp. 8–32.

The conclusion is reached on the basis of the above analysis, that (*a*) forgetting of factual material read a single time follows the same general law of rapid initial loss and more gradual subsequent loss as was found by other investigators for rote memory; (*b*) forgetting of factual material proceeds more slowly than forgetting of nonsense syllables; (*c*) forgetting of factual material proceeds more slowly than forgetting of meaningful material learned by rote; and (*d*) it takes much longer to approximate a point of complete forgetting for factual material than it does either for nonsense syllables or meaningful material learned by rote. It should be remembered that the above comparisons are between memory for facts apprehended in a single reading of an article as measured by a recognition test and verbatim memory for materials learned by rote to the point of one perfect and unhesitating reproduction.

Correlations between Memory Intervals

Introduction.—The problem of forgetting will be still further clarified by an analysis of the relationship existing between the scores of the same individuals after different intervals,—*i.e.*, of the question, "Do pupils rank in the same order, in memory after different time intervals?" This can be done most expeditiously by the correlation method.

Several workers have investigated the correlation between immediate and delayed memory. Among these, Thorndike found a correlation of .9 between immediate recall for twelve unconnected words and delayed recall after twenty-four hours.[1] Gates found correlations of .73 to .89 between immediate and delayed recall of nonsense syllables and of biographies. His subjects were children in the elementary school.[2] Bassett reports average correlations of .838 between immediate recall of history courses and recall after four months, .797 between immediate recall and recall after eight months, .839 between immediate recall and recall after twelve months, .840 between immediate recall and recall after sixteen months, .935 between recall after four and eight months, .915 between recall after eight and twelve months, and .961 between recall after twelve and sixteen months.

[1] Thorndike, E. L.: The Relation between Memory for Words and Memory for Numbers, and the Relation between Memory over Short and Memory over Long Intervals. *American Journal of Psychology*, Vol. XXI, 1901, pp. 487–488.

[2] Gates, A. I.: Correlations of Immediate and Delayed Recall. *Journal of Experimental Psychology*, Vol. IX, 1918, pp. 489–496.

She concludes that time and forgetting do not greatly reduce the relative standing of pupils in retention of history learned in school.[1]

The correlations obtained by the authors cited are, in general, fairly high. Indeed, Thorndike makes the statement, "The relation

TABLE XI.—CORRELATIONS BETWEEN FACTUAL MEMORY AFTER VARIOUS INTERVALS

V, Variables	*r*	PE	*r* corrected for attenuation	*N*
Immediate Radium; one day Germans	.75	.017	.85	293
Immediate Germans; one day Arkwright	.78	.016	.83	284
Immediate Arkwright; one day Radium	.73	.019	.80	286
Average: Immediate; one day	.75	....	.83	
Immediate Radium; fourteen days Germans	.61	.026	.66	273
Immediate Germans; fourteen days Arkwright	.50	.030	.54	289
Immediate Arkwright; fourteen days Radium	.53	.029	.58	282
Average: Immediate; fourteen days	.55	....	.59	
Immediate Radium; thirty days Arkwright	.48	.032	.53	268
Immediate Germans; thirty days Radium	.50	.030	.54	288
Immediate Arkwright; 30 days Germans	.58	.030	.62	225
Average: Immediate; thirty days	.52	....	.56	
Immediate Radium; one hundred days Arkwright	.36	.039	.39	223
Immediate Germans; one hundred days Radium	.48	.032	.52	261
Immediate Arkwright; one hundred days Germans	.50	.036	.54	204
Average: Immediate; one hundred days	.45	....	.48	
Fourteen days Radium; thirty days Germans	.58	.032	...	192
Fourteen days Germans; thirty days Arkwright	.58	.030	...	223
Fourteen days Arkwright; thirty days Radium	.50	.033	...	241
Average: fourteen days; thirty days	.55	....	...	
One day Radium; one hundred days Germans	.33	.056	...	115
One day Germans; one hundred days Arkwright	.33	.062	...	93
One day Arkwright; one hundred days Radium	.38	.053	...	119
Average: one day; one hundred days	.35	....	...	

between retention of the effects of an experience for one or two minutes and their retention for one or two days thus seems to be one of the closest yet measured in human nature." However, we must remem-

[1] Bassett, S. J.: "Retention of History in the Sixth, Seventh and Eighth Grades with Special Reference to the Factors That Influence Retention." Johns Hopkins Studies in Education No. 12. The Johns Hopkins Press, Baltimore, 1928, pp. 22–24.

ber that a correlation of as much as .90 reduces the standard error of estimate to only a little less than the standard deviation.

Result of Present Study.—The correlations found between the various memory intervals of the present study are shown in Table XI. Each correlation is based upon the largest number of cases available for the purpose. Since different materials were used at the intervals compared, the coefficients are attenuated. Corrections have been made for attentuation, and these are presented in a separate column. Without doubt the corrected coefficients are much nearer the truth than the raw coefficients; however, even they are far from perfect.

The correlations here reported indicate a fair degree of positive correlation between the various intervals. A study of the table reveals an interesting fact,—*viz.*, that the correlation coefficients become smaller the greater the interval between the memories measures. The averages of the correlations are in the following order:

Variables	Time intervening	Average correlation
Immediate; one day	one day	.75
Immediate; fourteen days	fourteen days	.55
Fourteen days; thirty days	sixteen days	.55
Immediate; thirty days	thirty days	.52
Immediate; one hundred days	one hundred days	.45
One day; one hundred days	ninety-nine days	.35

It seems, from these facts, that the relationship existing between memory at different intervals is influenced the more by uncorrelated factors as time goes on. This problem deserves further investigation. The writers offer the hypothesis that, among other factors, guessing plays a larger and larger rôle with the increase of time interval between reading and testing. This conjecture, indeed, is supported by the fact revealed by an examination of the test papers, that the subjects make fewer attempts the longer the intervals, indicating that they are more certain of their answers in the shorter periods than in the longer ones.

Summary of Results

1. Immediate and delayed factual memory for an article read a single time varies with the material used, the more familiar articles,

Radium and Germans, being remembered better than the less familiar Arkwright.

2. Negative skewed distributions in immediate and one day delayed recall for Radium and Germans indicate a piling up of scores in the range above the mean; Arkwright gives normal distributions at these intervals.

3. Wide individual differences, extending over about three-fourths of the scoring range, are present in the abilities of subjects to remember what they read a single time.

4. Memory for periods longer than one day result in slightly positively skewed distributions.

5. Immediate memory for an article read a single time increases with grade from fifty per cent in the seventh to sixty-five per cent in the twelfth for the material "Arkwright," and similarly for other materials.

6. One day delayed memory for Arkwright likewise increases with grade from forty per cent in the seventh to sixty-one per cent in the twelfth.

7. No marked relationship exists between grade and factual memory over fourteen days and longer.

8. Factual material is forgotten similarly to nonsense syllables and meaningful materials learned by rote,—*viz.*, rapidly at first, then more and more slowly as the interval between learning and the measurement of memory increases.

9. When memory is based on the amount of the article actually apprehended by the pupils, the initial forgetting for factual material is less abrupt than for nonsense syllables.

10. Factual material is forgotten less rapidly than either nonsense syllables or meaningful material learned by rote.

11. The point of complete forgetting is not nearly approximated even after one hundred days.

12. An appreciable amount (six per cent for Arkwright) is forgotten during the time of reading and writing the test.

13. Positive correlation exists between immediate memory and delayed memory, also between delayed memory at different intervals.

14. The degree of correlation existing between memory at different intervals vary inversely with the length of the interval. They range from an average of .75 between immediate and one day delayed memory to .35 between one day and one hundred day delayed memory.

Comments

1. Since the type of material which pupils read and study markedly influences the amount which they apprehend and remember, methods should be devised for scientifically matching the difficulty of textbooks to the ability level of the pupils for whom they are intended.

2. More research is needed to determine the factors inherent in material read and in the mental make-up of pupils which determine the efficiency of memory.

3. It is obvious that a single reading of lesson assignments is an inefficient method of study. Studies are needed to discover more effective study techniques. The influence of repeated readings, summarizing, reading to find answers to questions and to solve problems, are suggested problems. A veritable mine of similar problems are suggested throughout the pages of G. A. Yoakam's recent book, *Reading and Study*.[1]

4. Other interesting fields of experimentation are suggested by statements nine, ten, eleven and fourteen of the foregoing summary.

---

[1] The Macmillan Co., New York, 1928.

EFFECT OF ORDER OF PRESENTATION ON THE RECALL OF PICTURES

DANIEL D. DROBA

Ohio State University

Results described in this report were obtained in connection with an experiment on the effect of printed information on memory for pictures.[1] One hundred twenty students attending the University of Chicago were used for subjects. For memory material twenty uncolored postcard pictures of paintings in the Chicago Art Institute were used throughout as follows:

1. A Woman in Gray, by William Orpen, English, 1878– (14)
2. The Fountain, by John S. Sargent, American, 1856–1925. (16)
3. Joan of Arc at the Court of Chinon, by Maurice Boutet De Monvel, French, 1850–1913. (15)
4. Arab Scouts, by Adolph Schreyer, German, 1828–1899. (20)
5. The Music Lesson, by Gerard Terberg, Dutch, 1617–1640. (19)
6. The Young Duchess, by John Faed, Scotch, 1820–1902. (10)
7. Marches in the North of Holland, by Eugene Jettel, Austrian. 1835–1901. (5)
8. The Potato Harvest, by Ludwig Knaus, German, 1829–1910. (11.5)
9. Democritus, The Laughing Philosopher, by Jusepe Ribera, 1588–1656. (17)
10. Portrait of Helena Dubois, by Anthony Van Dyck, Flemish, 1599–1641. (18)
11. Rembrandt's Father, by Rembrandt Harmensz Van Rijn, Dutch, 1606–1669. (2.5)
12. The Wrestler's Challenge, by Mihaly Munkacsy, Hungarian, 1846–1900. (11.5)
13. Night in the Garden of Gethsemane, by Louis Cranach, German, 1472–1553. (7)
14. A Wrangle Over Cards, by Jean David Col, Belgian, 1822–1900. (1)
15. The Assumption of the Virgin, by Domenikos Theotokopoulos, Spanish, 1547–1614. (4)
16. The Repentance of Simon Peter, by Jean Charles Cazin, French, 1841–1901. (6)
17. Pilots, Puerto De Los Pasajes, Spain, by Frank W. Brangwyn, English, 1867– (13)
18. He That Is Without Sin Among You, by Benjamin West, American, 1738–1820. (9)
19. The Water Mill, by Meindert Hobbema, Dutch, 1628–1709. (8)
20. In Holland Waters, by Paul Jean Charles Clays, Belgian, 1819–1900. (2.5)

---

[1] D. D. Droba: Effect of Printed Information on Memory for Pictures. *Museum News*, Sept., 1929; *Psych. Abstracts*, No. 94, 1930.

Each picture was fastened on a white card and the information pertaining to the picture was typed either at the bottom or to the right of the picture. Six types of information were used ranging from no information to an information including the title of the picture, the last name of the painter, and a sentence about the painter.

The pictures were shown individually to every subject, the exposure time being fifteen seconds for every picture. Immediately after all the pictures were presented the subject was asked to describe on a separate card all the pictures he was able to recall. Twenty subjects were used for each of the six conditions of information.

For studying the relation between the order of presentation and frequency of recall two methods were used. The first was the rank correlation method. Hence, we ascertained the number of times a picture was recalled by the one hundred twenty subjects. The pictures were then ranked on the basis of the frequency of recall. The most frequently recalled picture was put first, the next most frequently remembered picture was put the second, and so on (rank numbers of recall are given in the parenthesis after each picture in the list). The rank correlation between the frequency of recall and the order of presentation was found to be $-.61$. This negative correlation tends to show that the earlier a picture is presented the less likely it will be recalled.

The second method used was the graphic method. The twenty pictures were divided into two groups: The ten pictures presented first and the ten pictures presented second. The averages of the number of pictures recalled under each of the six conditions of information was calculated. These averages, together with the total averages for all the one hundred twenty subjects, were tabulated in Table 1.

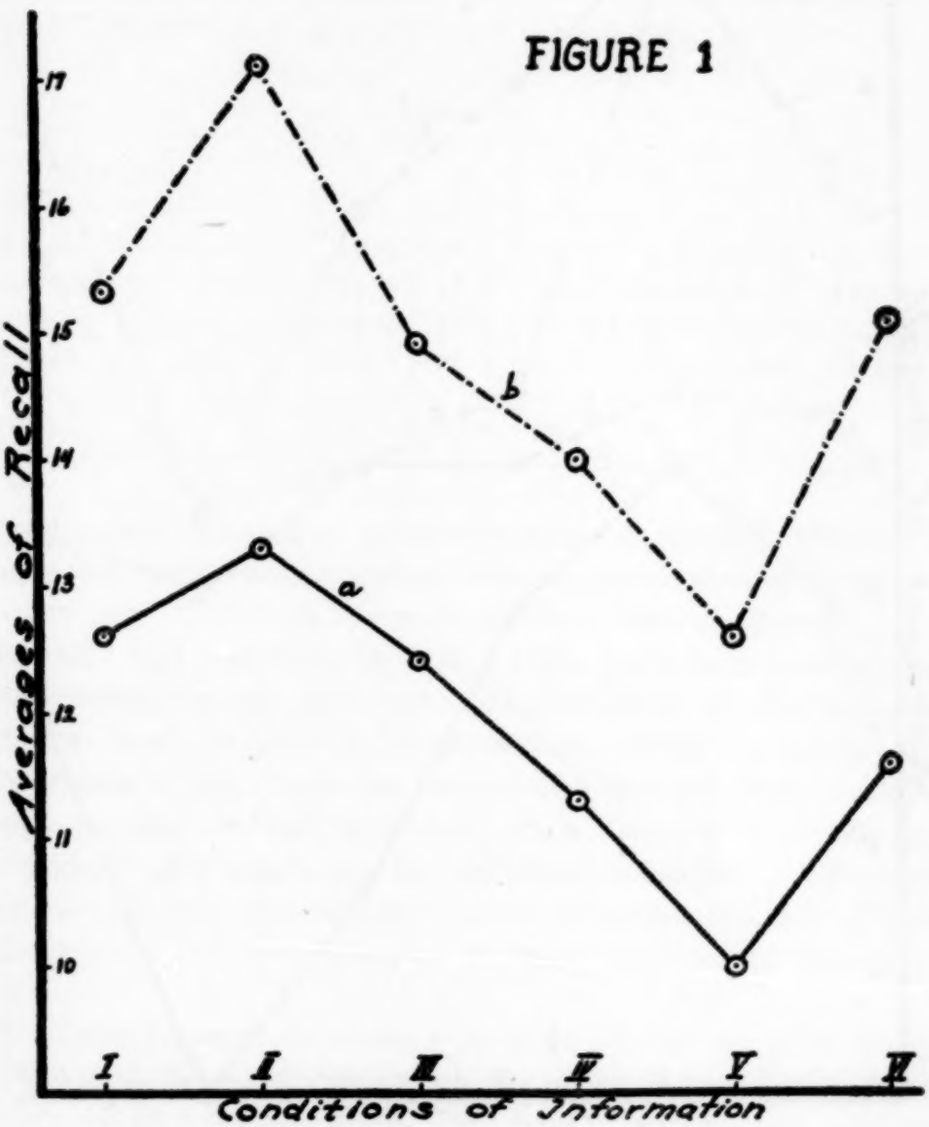
TABLE I

Pictures	I	II	III	IV	V	VI	Total
a. First ten	12.6	13.3	12.4	11.3	10.0	11.6	11.8
b. Second ten	15.3	17.1	14.9	14.0	12.6	15.1	14.8
Subjects	20	20	20	20	20	20	120

Under each condition of information, designated by Roman numbers from I to VI, a higher average was found for the second half of the series of pictures. The difference between the total averages

for all the one hundred twenty subjects is 3.0 which is a confirmation of the above finding that pictures presented in the second half of the series were recalled more frequently, on the average, then pictures presented in the first half of the series.

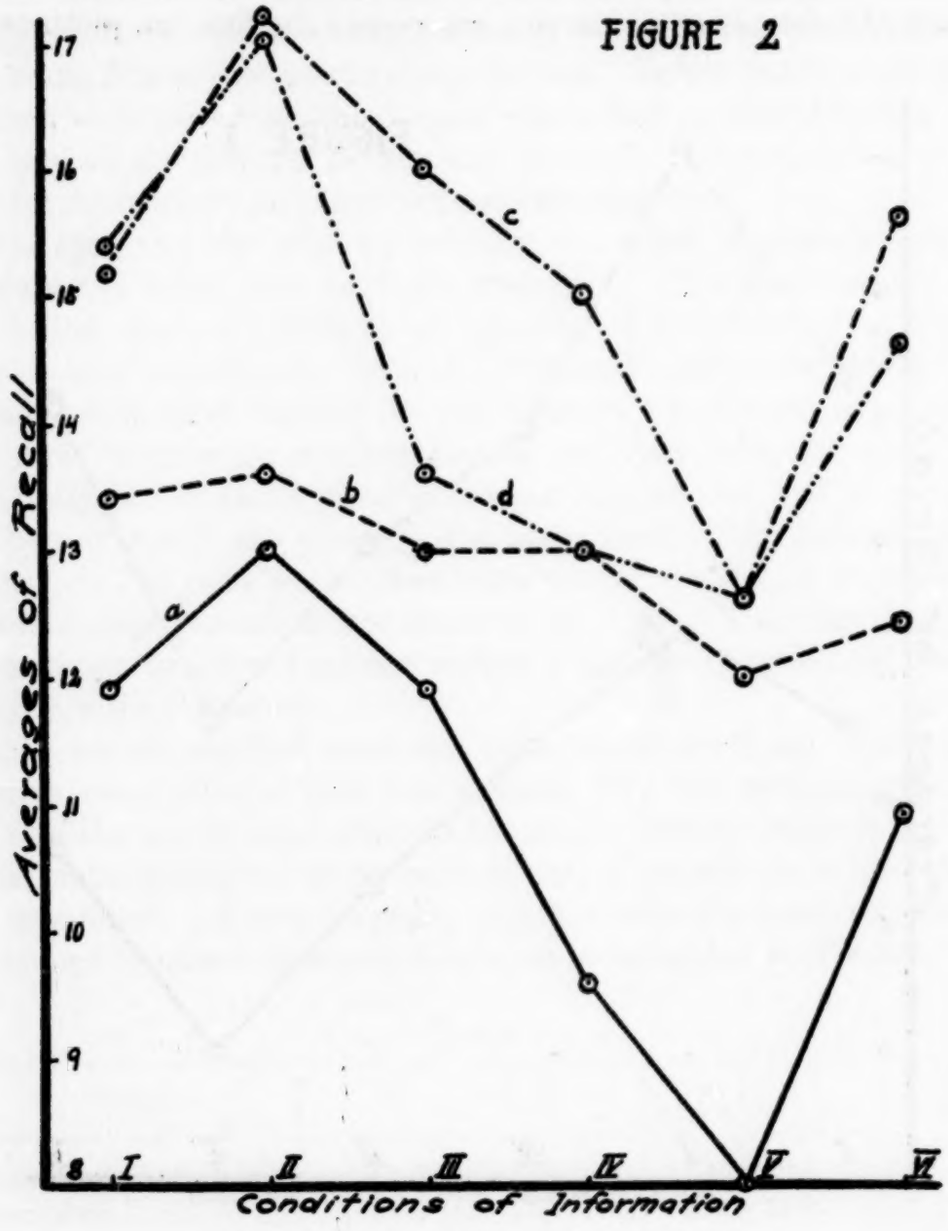
Figure 1 graphically represents these averages plotted against the amount of information. Curve *a* represents the first ten pictures and



curve *b* the second ten pictures. The difference between the two curves is readily seen: *b* is much higher in all conditions than curve *a*. This means that pictures located in the second half of the series were consistently recalled more times than those in the first half of the series.

The two halves of the pictures were then again divided into two equal subgroups so that we obtained four groups of five pictures each. Similarly, the averages of recall were calculated for each group under each condition, and tabulated in Table II.

FIGURE 2



The total averages for all the one hundred twenty subjects are also included in the table. According to the total averages (the last column) pictures in the third quarter were recalled the most frequently. Pictures located in the last quarter come next, the second group the next, and the first group is the last of all.

Data shown in Table II were plotted in Fig. 2. The highest curve is curve *c* representing the third group of five pictures which indicates that the best position is the third quarter in all except the first and the fifth conditions. In the latter condition it is equal with the last quarter. Curve *d* representing the last quarter drops down especially in the medium conditions. Curve *b* representing the second quarter is different from the first two especially under the end conditions.

TABLE II

Pictures	I	II	III	IV	V	VI	Total
a. First five	11.9	13.0	11.9	9.6	8.0	10.9	10.8
b. Second five	13.4	13.6	13.0	13.0	12.0	12.4	12.9
c. Third five	15.2	17.2	16.0	15.0	12.6	15.6	15.2
d. Fourth five	15.4	17.0	13.6	13.0	12.6	14.6	14.3
Subjects	20	20	20	20	20	20	120

Pictures located in the first quarter and symbolized by curve *a* were decidedly less frequently recalled than pictures placed in the other three quarters. This fact is clearly evident from the figure.

An additional fact may be noted from both figures which is not strictly related to our problem. The relation of the amount of information and frequency of immediate recall is consistent for all the groups of pictures: the halves and the quarters. All curves first rise to the second condition, then they drop down to the fifth condition, and again rise for the sixth condition. The meaning of this relation was substantiated in a separate article.[1] However, the consistency of the relation is being mentioned here for the first time.

The following conclusions may be drawn from this experiment:

1. The rank correlation between the order of presentation and the frequency of immediate recall seems to indicate that priority of presentation is not conducive to such recall.

2. In all the six conditions of information the position of a picture in the second half of the series is more favorable to immediate recall than in the first half.

[1] *Ibid.*

3. Pictures located in the third quarter of the series were recalled the most frequently of all.

4. Position in the first quarter is decidedly less favorable to immediate recall than positions in the other three quarters.

5. Effect of information on memory for pictures, described in a previous article, was found to be consistent for all the above mentioned positions.

THE INFLUENCE OF AN AUDIENCE UPON RECALL

CLARA BURRI

Graduate Student, University of Chicago

The Problem

Introspective reports from many observers indicate that the presence of an audience frequently has a detrimental effect upon performance. Many individuals relate that the presence of others causes them many difficulties of behavior especially in speech and mnemonic responses.

Sometimes we find that a student comes to school, seemingly unprepared, yet he maintaines that: "He did learn his lesson, but that he just could not remember when others were watching." Every teacher is familiar with this or similar statements and he frequently considers it as a mere excuse for not having studied. This investigation is an attempt to study experimentally some of the fundamental phases of the above problem. Our task is to discover whether an individual who has learned certain memory material, and is then required to recall that material before a group of people, will as a consequence of the presence of that audience, be able to recall as readily as when he is alone.

Few people have considered the influence of a group on an individual's efficiency to recall when learning did not take place in that group. Elkine[1] is the only author who reports any data concerning group influence on recall or reproduction. He reports that memory is better in the group. Forty children were given one presentation of a series of one-syllable words and were then called upon for a written reproduction immediately after the presentation, and another one day later. First, the individuals went through the above stated procedure alone, and later as members of a group. Elkine explains his results in terms of "rivalry and the influence of suggestibility due to contact and mere presence of other people." Elkine's study differs from the one reported in this paper in one important aspect. He studied the influence of a group on learning and memory when the subject was a member of the group. The results obtained may therefore be explained as being due to rivalry. In this study particular

[1] Elkine, E.: Influence de group sur la memoire. *J. de Psychol.*, 1927, pp: 827–830.

care was taken to eliminate, as far as possible, the factor of competition, and hence the subjects were tested when recalling before people who were not engaged in any comparable activity.

Method and Procedure

Subjects.—Sixty students, nineteen men and forty-one women, from the classes in introductory psychology, and from other departments of the University of Chicago, participated in the experiment.

Conditions.—The subjects were divided at random into three groups of twenty each. Group one acted as control, learning and recalling under the normal condition, with only the experimenter present. Group two learned under the same condition as did group one, but they recalled and relearned before an audience consisting of four persons who sat in the experimental room without paying any attention to them. The individuals of group three also learned alone, and then recalled and relearned before an audience. This audience however attentively watched and kept track of their responses.

Audiences.—The audiences consisted of two men and two women, all graduate students or majors in psychology. They were selected with a view to their prestige in the eyes of the subjects, and as far as possible consisted of people who were not acquainted with the subjects. No attempt was made to keep the same individuals for spectators throughout the experiment. As has already been suggested two kinds of audiences were employed: One in which the members paid no attention to the subject, but were merely present in the experimental room: Another, an attentive audience, the members of which listened to the subject and checked his responses on a slip of paper which contained the lists of words in their correct order.

Material.—The material to be learned consisted of a list of fifteen pairs of words which had been selected at random from a standard dictionary. In constructing the list care was taken that no letter should occur twice in a given pair. The words were exposed by means of a hand-operated memory drum, a standard instrument in the laboratory of the University of Chicago. Other materials used were cards three to five inches with the response words copied on incorresponding order with the stimulus words of the test series, for recording the responses. A stopwatch served for timing the exposure of the paired associates.

Procedure.—The paired associate list was presented once, allowing three seconds for each exposure, after which the subject was tested for

recall. In the test series each stimulus word was presented alone and in a different order from that in which it originally appeared, and the subject had to give the proper response word within the given time of three seconds. This entire procedure was continued until the subject was able to respond correctly to every stimulus word in the list. The criterion of learning was one errorless recall. A subject learned the list of paired associates and then recalled and relearned them twenty-four hours later either before the experimenter alone, or in presence of an audience, attentive or inattentive, depending upon the particular condition desired.

Instructions.—At the beginning of the experiment each subject was required to read a carefully prepared and typewritten instruction sheet, after which he was quizzed briefly to make certain that he understood the general nature of his task. At the same time he was also told under what condition he was expected to recall, and relearn on the following day. Thus every subject knew before he started to learn, whether an audience, and if so, what kind of an audience, was to be present for recall.

The only subjects given other instructions on the recall day were those who were to face an attentive audience. To them, the request was made to "respond distinctly and loudly, so that every person in the room can understand." This extra order was given to enhance the effectiveness of the attentive audience and thus to further increase the difference between the non-attentive and attentive audience.

Treatment of Data.—The amount of retention is stated in terms of (1) number of words recalled; (2) number of trials required to relearn; and (3) the saving score, which was obtained by the following formula:

$$1 - \left(\frac{\text{Trials to relearn}}{\text{Trials to learn}}\right)100$$

The difference in efficiency of recall between the several conditions is determined statistically and the results are stated in terms of the numerical differences, the averages, the probable errors of the means, the difference between the means, and the probable error of their differences.

Results

Table I gives the total and mean number of words recalled, trials required to relearn and the saving score for the different conditions. If we analyze the data it appears that the results in terms of recall are most clear cut, while those in terms of relearning and saving, although

less large, still show a definite tendency. The efficiency of retention as measured in terms of recall, relearning, and saving, is greatest for no audience. In the case of an audience, and an attentive audience, the numerical scores for retention are considerably smaller than those for no audience, while on the other hand they are about equal for the two kinds of audiences. The statistical significance of the findings is shown in the later tables. These results throw light upon a fundamental principle of reproduction. They indicate that any stimulus-

TABLE I.—THE EFFICIENCY OF RETENTION IN THE DIFFERENT CONDITIONS

Condition	Recall; syllables recalled	Relearning; trials to relearn	Saving; percentage of learning time saved
No audience.			
Total	269	18	1561
Average	13.45	.90	78.05
Audience.			
Total	237	33	1387
Average	11.85	1.65	69.35
Attentive audience.			
Total	234	32	1417
Average	11.70	1.60	70.85

response situation must be considered as a functional whole consisting of a constellation of numerous organismic and environmental factors. Any change in the situation will affect the synthesis of the whole. If a behavior pattern is once formed it may be reinstated best in a similar situation to that in which it was originally acquired. In our problem not only the total organization present at the time of learning, but the constellation of organismic and environmental factors at the time of recall were equally important in facilitating or inhibiting reproduction.

Table II shows the mean number of words recalled, of trials required to relearn, and of the saving per list for each of the two conditions, no audience and audience; the probable errors of the two means from the three criteria and the probable errors of their differences. There is a difference between the two conditions in all of the three criteria. The one in terms of recall is almost four times its probable error. The differences in the case of the relearning and

saving scores is not so great, yet they indicate a consistent tendency. This is especially remarkable since these scores are not direct measures of recall.

TABLE II.—THE DIFFERENCE BETWEEN NO AUDIENCE AND INATTENTIVE AUDIENCE

Condition	Recall; syllables recalled	Relearning; trials to relearn	Saving; percentage of learning time saved
No audience.			
Mean	13.45	.90	78.05
PE	.28	.12	4.14
Audience.			
Mean	11.85	1.65	69.35
PE	.33	.21	4.67
Difference between the two means	1.60	.75	8.70
PE	.43	.24	6.24

In Table III we find the mean number of the words recalled, of trials required to relearn, and of the saving per list for each of the two conditions, no audience and attentive audience, the probable errors of the two means from the three criteria, and the probable error of their differences.

TABLE III.—THE DIFFERENCE BETWEEN NO AUDIENCE AND ATTENTIVE AUDIENCE

Condition	Recall; syllables recalled	Relearning; trials to relearn	Saving; percentage of learning time saved
No audience.			
Mean	13.45	.90	78.05
PE	.28	.12	4.14
Attentive audience.			
Mean	11.70	1.60	70.85
PE	.17	.16	4.23
Difference between the two means	1.75	.70	7.20
PE	.32	.20	5.91

As before we find that the subjects with audience give poorer results. The difference between no audience and attentive audience

is even larger than is the difference between no audience and inattentive audience. Again we get the largest divergency for recall with a disparity between the two averages of 1.75 and a probable error of the difference of .32, which is less than one-fifth of the actual difference. This indicates a high degree of significance.

The results become more striking when we consider that the differences exist in spite of the fact that the subjects took, on an average, longer to learn the paired associates when they were anticipating an audience, than when expecting to recall only before the experimenter. We generally find that a longer learning period is followed by better recall. One can scarcely avoid the inference that this lengthened learning period is an outcome of the expectancy of an audience. The discussion of this topic, however, will have to be left until later.

TABLE IV.—THE DIFFERENCE BETWEEN INATTENTIVE AUDIENCE AND ATTENTIVE AUDIENCE

Condition	Recall; syllables recalled	Relearning; trials to relearn	Saving; percentage of learning time saved
Audience.			
Mean	11.85	1.65	69.35
PE	.33	.21	4.67
Attentive audience.			
Mean	11.70	1.60	70.85
PE	.17	.16	4.23
Difference between the two means	.15	.05	1.60
PE	.37	.26	6.30

Table IV gives the mean number of words recalled, of trials required to relearn and of the saving per list for each of the two conditions, inattentive and attentive audience; the probable errors of the two means from the three criteria and the probable errors of their differences. There is practically no disparity between the results for the two kinds of audiences.

It seems that the kind of audience is not very important as regards its influence on recall. However, our results may be a function of the particular situation, and if that should be the case, other group arrangements might have different consequences. Therefore no general conclusions can be made. From the viewpoint of the organization of the audiences further observations and studies would be desirable.

Table V contains the data on the learning of the paired associates. In all cases learning took place with only the experimenter present. The sole difference between the several learning conditions was that, in the first, the subjects did not expect an audience, while in the other two, they did. Naturally we would expect the learning efficiency to be about equal throughout the experiment, or at least not to vary to a greater degree than could be explained as due to chance. Yet we find that learning was on the average considerably easier when the subjects did not expect an audience. The difference in trials required to learn the series of paired associates when expecting no audience and when anticipating an audience is a total of 28 trials or a mean of 1.40

TABLE V.—THE EFFECT OF EXPECTANCY OF AN AUDIENCE FOR RECALL, AT THE TIME OF LEARNING, ON THE RATE OF LEARNING

Condition	Total trials to learn	Average trials per list	PE
No audience	96	4.80	.32
Inattentive audience	124	6.20	.32
Attentive audience	123	6.15	.34
Difference between no audience and attentive audience	28	1.40	.45
Difference between inattentive and attentive audience	0	0	

trials. The probable error of the difference of the two means is .45. The actual disparity is a little over three times its probable error, indicating significance. There are no significant differences for the two different kinds of audiences.

The fact that the people needed more trials to learn the material when they were expecting an audience may be explained as being due to difference in anticipation, and integration with reference to the learning situation. Anticipating an audience seemed to have been inhibitive to the learning process. The subjects' reports clearly indicate that their knowledge of having to recall before spectators caused some uneasiness. They appeared to learn with the intention of knowing the words well, and of being able to tell them to somebody. Such an attitude was not recognizable with the subjects who did not look forward to an audience; these subjects appeared to learn without any definite end in view. These are the facts as they are before us. We

conclude from them that differences in expected conditions of recall may cause corresponding differences in rate of learning.

Conclusions

Restating the facts as they appear from our investigation we find that:

1. If learning takes place privately, the presence of a group at the time of recall is detrimental to the efficiency of reproduction.

2. The kind of audience is not very important as regards its influence on recall. However, it may be that in this connection the results are a function of the particular situation, and that other group arrangements might have different consequences.

3. The anticipation of an audience for recall, at the time of learning, increases the learning time, and, in spite of this, the amount remembered is less when recall occurs before an audience.

These results suggest that the pupil who complains of not being able to remember well when requested to recite before a class may not be trying to escape responsibility. His inability to recall reveals a fundamental factor of memory, and shows that the situation during the time of learning and during the time of recall are equally important in facilitating or inhibiting the response.

THE RELATIVE INFLUENCE OF VISUAL AND AUDITORY FACTORS IN SPELLING ABILITY

GEORGE W. HARTMANN

Department of Psychology, Pennsylvania State College

A. Background of the Problem

The spelling of the average undergraduate is notorious. One suspects that the great vogue of objective examinations may be partly explained by their power to alleviate eyestrain due to illegible papers, disgust caused by childish syntax, and irritation born of repeated misspellings. These common deficiencies are especially prominent in science courses where the difficulties of mastering a technical vocabulary appall even the superior student. How many instructors are there who after exposing their charges to a year of its content find them spelling psychology more like philosophy, physiology, and philology than like its own true self?

The various explanations which have been advanced for special abilities and disabilities as they manifest themselves in academic or vocational fields are suggestive but hardly convincing. In the case of such a restricted function as spelling, *e.g.*, it is easy to demonstrate that it is not an exclusive property of high or low intelligence, although it is equally certain that it is not as independent of general native ability as some extremists would have us believe. An analysis of the commonest types of errors leads to the suspicion that various sensory and motor idiosyncrasies are the major determiners of performance, but *which* ones are crucial is as yet unknown. The current view seems to be that every poor speller must be individually diagnosed and specific remedies prescribed—an obvious consequence of an implicit belief in plural causation.

However, the scientific search for adequate generalizations is not to be dismissed so lightly. Just as it is profitable to know that superior intellects tend to have correspondingly better physiques, emotional control, etc., so we ought to know what the significant characteristics of a good speller are as distinct from a bad one.

Among the many explanations advanced for poor spelling, perhaps the most recurrent one is that which attributes it to the non-phonetic nature of the English language. Were certain sounds uniformly

represented by definite letter-forms this alone would facilitate correct spelling. The advocates of reformed spelling apparently take this seriously, despite the fact that errors occur with relatively equal frequency in highly phonetic tongues such as German or Spanish. A slightly more plausible variant of this argument is that if a person could pronounce words correctly he would have no difficulty with their orthography. Unfortunately, this fails to account for the fact that many Americans can spell French terms properly but pronounce them in a very unorthodox manner; and similarly for the Frenchman using English. Moreover, most misspellings are substitutions or displacements of letters which by no stretch of imagination can be attributed to faulty pronounciation. The definitive elimination of the pronunciation theory would seem to be inevitable consequence of the finding that deaf-mutes are relatively better in spelling than in other school subjects.

Other viewpoints have emphasized the rôle of visual perception. Here the commonest opinion is that poor spellers are deficient in visual imagery, but the tachistoscopic span of apprehension has also been considered as the decisive factor.

Where such diverse and contradictory opinions exist, a resort to experimental attack is indicated. To the writer, the problem phrased itself as follows:

Does excellence of auditory or visual perceptive capacity play the dominant rôle in spelling performance? Is spelling ability more closely associated with those reactions mediated by the ear or with those involving the eye? The wording of these questions by implication minimizes the influence of motor impulsions as it is assumed that chance slips of the pen should occur equally often with both groups; and it also sets in the background the influence of such central factors as intelligence, type of imagery, association, etc.

B. The Testing Procedure

This investigation began by administering to large samples of each college class from Freshman to Senior (six hundred thirty-six cases in all) a test consisting of fifty selected words appearing in both the Thorndike and Horn Word Books. Upon a prepared sheet, the subjects *printed* each word as it was spoken and its meaning illustrated by a sample sentence. That the test was a reasonable measure of spelling ability is indicated by the reliability coefficient of .84 $\pm$ .007, obtained by the split-half technique. For the purposes of this experi-

ment, representative cases chosen from the middle and both tails of the distribution were selected for intensive study. Sixty-three individuals were thus obtained—twenty-four of the best spellers, eighteen of the poorest, and twenty-one average ones centering around the median score.

The following test series was administered individually to each of the subjects:

Test 1. *Perceptual Span for Meaningful Material* (*Reliability* .74).—Upon the middle of twenty small white cards were printed the following unusual words, derived largely from Whipple's well-known information test:

ageratum	simony
Caedmon	synecdoche
cleistogamous	trephine
gneiss	trilobite
guimpe	Weismannism
hemiptera	onomatopoeia
impetigo	sphygmomanometer
intaglio	exophthalmic
metacarpal	Abderhalden
peneplain	La Rouchefoucauld

One-second exposures of the cards were made tachistoscopically; during the intervals (which varied with the writing time) the subject recorded the word he had seen. The total number of reproductions correct in every detail constituted the score.

Test 2. *Visual Recognition* (*Reliability* .64).—Twenty common words printed on cards were exposed to the subject at the rate of about one per second with instructions to fixate each item. Immediately after the original series had been seen, the cards were shuffled with twenty other confusion stimuli. The mixed set was then arranged by the subject into two piles—one containing the cards the subject had seen and the other containing those he had not seen. Score = 2 × number of errors in both piles subtracted from 40. (For the exact items see the list in the writer's appendix to *Archives of Psychology*, No. 100.)

Test 3. *Silent Reading*.—Whipple's High School and College Form B was administered in accordance with the time limits and other conditions specified in his directions sheet. Score equals the number of correct answers.

Test 4. *Hidden Word Identification* (*Reliability* .88).—This consists in underlining a meaningful English word in a page of pied material, the material being embedded in a jumbled nonsense collection. A two-minute time limit was imposed. Score equals the number of words properly underlined.

Test 5. *Letter-digit Substitution Code* (*Reliability* .56).—A simple "home-made" code using the ten Arabic figures was devised. Time of transcription was measured in seconds.

Test 6. *Auditory Memory Span for Digits* (*Reliability* .73).—The fourteen number sets from the corresponding section of Whipple's manual were read to the subject in time with a metronome, and recorded by him as soon as heard. Averaging the highest correct reproduction in both sets gave the desired measure of span.

Test 7. *Pronunciation* (*Reliability* .75).—The thirty-two words forming the pronunciation section of the Cross English Test were used. The printed list was shown the subject with instructions to pronounce each word as though he were being graded for purity of speech by a language teacher. The experimenter checked the correct items by listening carefully as the subject proceeded.

Test 8. *Auditory Recognition* (*Reliability* .53).—A list of ten ordinary words was read at an even rate to the subject with instructions to fixate them. Immediately after this list had been read, a longer list of twenty words containing the ten originals plus ten confusion stimuli, was offered. As each word in the test list was uttered, the subject said "yes" if he had heard it before and "no" if he had not.

The reliability coefficients presented for each of the foregoing measures were obtained by duplicate forms in the case of Tests 2, 4, 5, 8 and by matching alternate items in tests 1, 6, 7; for Test 3 adequate reliability standards were assumed to have been met by the originator.

C. Results

The mean performance on each of the preceding eight tests was separately computed for each group of spellers—the superior, the average, and the inferior. Table I below summarizes the essential data for all classes.

Before proceeding to an interpretation of the table, let it be noted that if spelling ability be conditioned by excellence in visual perception then the "good" group should reveal a definite superiority in Tests

TABLE I.—COMPARATIVE PERFORMANCE OF THREE CLASSES OF SPELLING ABILITY IN VARIOUS VISUAL AND AUDITORY MEASURES

Spelling scores	1 Meaningful perceptual span	2 Visual recognition	3 Silent reading	4 Hidden words	5 Substitution	6 Auditory memory span	7 Pronunciation	8 Auditory recognition	Constants	Class
38.25	9.79	69.46	9.34	13.76	61.01	6.96	17.55	82.65	Mean	Total group
7.32	3.07	13.20	3.06	3.90	8.86	1.88	2.61	7.78	SD	
47.62	12.46	75.63	10.96	14.85	55.52	7.15	19.96	85.94	Mean	
1.26	1.95	11.87	2.82	3.39	6.42	.75	3.75	7.69	SD	Superior
29.15	4.7	2.09	2.05	.88	3.81	.33	1.34	1.4	$D/SD_{diff.}$	
38.0	9.76	67.55	9.29	13.85	64.09	7.05	18.62	82.98	Mean	
1.95	1.89	13.79	2.62	4.18	8.48	1.22	2.91	6.43	SD	Average
18.24	4.81	1.15	2.43	1.29	.24	1.05	5.08	2.37	$D/SD_{diff.}$	
23.77	6.28	62.36	7.28	12.22	64.75	6.28	14.22	77.92	Mean	Inferior[1]
2.81	2.53	14.27	2.55	3.68	8.43	2.53	2.52	6.81	SD	

[1] The $D/SD_{diff.}$ comparison is of course unnecessary with the third group.

1–5 inclusive, but if excellence in auditory reactions be the decisive factor then the "good" group should be superior in Tests 6–8 inclusive. The expected positions of both the normal and poor groups under either hypothesis may be readily inferred.

The object of Table I is to show which tests meet the conditions of standard reliability in exposing the significant differences between groups representing three degrees of spelling ability. It will be seen at once that very few do so. In the visual series, Test 1 alone discriminates the superior from the average and the average in turn from the inferior. Tests 2, 3, and 4 all fall short of the conventional requirement. Test 5 oddly enough distinguishes the good from the mediocre group, but fails to discriminate between the latter and the inferior.[1]

In the auditory series, no test possesses discriminatory power with the exception of Test 7 which does differentiate the average from the poor spellers.

Apparently Test 1 of the visual group is the only one which can be used to segregate the various classes of spelling talent. This test contains a group of rare words which the subject copies from immediate memory as soon as they have been exposed. According to the evidence which this affords, speed and accuracy of immediate visual perception exert the dominant influence in spelling proficiency.

An attempt was next made to get at the heart of the problem by comparing total standard scores in both the visual and auditory batteries for each of the three ability groups. This comparison was made as follows: The means and sigmas of all subjects in each test were determined, and the mean sigma position of each ability group was separately computed for the visual and auditory series. This is a most laborious operation, but the following table gives the condensation:

TABLE II.—MEAN SD POSITIONS OF THREE SPELLING ABILITY GROUPS IN VISUAL AND AUDITORY TESTS

Visual battery (Tests 1–5 inclusive)			Auditory battery (Tests 6–8 inclusive)		
Superior	Average	Inferior	Superior	Average	Inferior
.55	−.06	−.62	.48	.04	−.69

[1] Since the scores in Test 5 represent times taken, the lowest values designate the best performances.

The reasoning which is intended to apply is: If the superior (or inferior) spellers are such primarily because of high (or low) visual perceptual capacity, then their relative performance should be most marked in the visual series. As a matter of fact, the order of excellence is about the same in both batteries. Good spellers, apparently, can use their ears proportionately as effectively as their eyes, which leaves us uncertain as to which of these senses their prowess should be attributed.

As the method of reliable differences yielded such ambiguous evidence, it was necessary to appeal to further analysis in the hope that this would reveal the true relations which conceivably had been obscured by some kind of compensatory masking. Raw correlations between the spelling scores and the eight tests were therefore computed, and the following coefficients obtained:

TABLE III.—RAW CORRELATIONS BETWEEN SPELLING SCORES AND PERCEPTUAL TESTS

Test No.	Description	*r*	PE
1	Meaningful perceptual span	.78	.03
2	Visual recognition	.39	.07
3	Silent reading	.47	.07
4	Hidden words	.27	.08
5	Substitution	.41	.07
6	Auditory memory span	.15	.08
7	Pronunciation	.58	.06
8	Auditory recognition	.43	.07

This method seems to check the results presented in Table I for again the test of perceptual span is predominant. Since the *r* between spelling and the data of Test 1 is attenuated by the imperfect reliabilities of the respective scores, Spearman's correction formula was applied, resulting in an *r* of almost .99! It appears that the two tests measure much the same function.

A final attempt was made to build up a regression equation from the three most satisfactory visual tests, *viz.*, Tests 1–3, inclusive. The raison d'être of this was that if it could be shown that admittedly visual factors possess adequate predictive power for degrees of spelling talent, then they must be viewed as causative influences thereof (in Hume's sense). The following calculations are accordingly presented:

$$R_{0(123)} = .784$$
$$x_0 = 1.89x_1 + .025x_2 + .19x_3 \text{ (deviation form)}$$
$$X_0 = 1.89X_1 + .025X_2 + .19X_3 + 16.24 \text{ (score form)}$$
$$SD_{(est.\ x_0)} = 4.54.$$

It will be noticed that the multiple correlation is in no way greater than the coefficient of zero order between the first test of our series and the spelling scores. The obvious implication is that this test carries almost exclusively the full burden of prediction.

D. Conclusions

The findings presented in the previous section do not readily lend themselves to any clear-cut interpretation, but certain provisional statements appear to be justified. In the first place, spelling ability is no more a function of *general* visual perception than it is of general auditory perception. Nevertheless it does seem to be closely related to the *special* reaction involved in reproducing tachistoscopically-exposed stimuli of a meaningful nature. To use the language of Gestalt, good spellers perceive total configurations of the verbal sort with greater facility than others. Secondly, spelling ability is essentially a central function; with normal subjects peripheral factors of an optical or retinal nature are inconsequential. Thirdly, while the best way to determine spelling talent is still the sample or scale method, it would seem possible to use the meaningful perceptual span technique as an alternative thereto, since this has the advantage of brevity. Fourthly, all the preceding statements should be interpreted in the light of the fact that all the data, especially the correlations, pertain to individuals chosen from each of the major segments of a larger curve.

E. Summary

This study concerned itself with the following question: Is there any form of sensory perception which is the major determiner of spelling proficiency? Sixty-three college students, representing respectively the best, the worst, and the middle levels of performance, were given eight laboratory tests *singly*. These tests fell into two groups—one requiring the use of the visual pathways and the other depending on the receptors for sound.

Statistical treatment of the results showed that the visual battery *as a whole* did not discriminate among the groups any better than the

general auditory battery, and *vice versa;* one visual test, however (that of immediate memory span for meaningful visual stimuli), correlated .78 ± .03 with the spelling criterion and served effectively to segregate the three classes of ability.

The major outcome of this investigation is that spelling ability is largely dependent upon *one special form* of visual reaction and not upon general superiority in any sense modality or upon a common integrative capacity.

FURTHER DATA CONCERNING THE EFFECT OF WEIGHTING EXERCISES IN NEW-TYPE EXAMINATIONS

C. W. ODELL

University of Illinois

The authors of several of the early standardized tests devoted considerable labor and care to the determination of supposedly accurate weights, usually on the basis of difficulty, for the separate elements composing their tests. Within a few years, however, a strong tendency to discontinue the use of such weights appeared. Studies by Douglass and Spencer[1] and others showed such high coefficients of correlation between scores based upon such weights and so-called "unweighted" or raw scores, that is, those based upon one point for each element, that they led to the conclusion that the considerable amount of extra work involved in deriving and using weights was unnecessary. In a recent article Corey[2] has presented some data which scarcely support this conclusion. He had each item of a new-type test in educational psychology weighted by six instructors in that subject. From the results he determined the correlations between raw scores and those computed according to these six series of weights, and also the effects of the different weightings upon letter marks into which the test scores were transmuted. Five of the six coefficients with the raw scores ranged from .82 to .88, the other being .96. Approximately one-fourth of the papers were given the same letter mark according to all seven scores, and another fourth each according to six, five, and four of the seven. The marks according to the six instructors varied from those based on raw scores in from twenty-two to forty-nine per cent of the cases.

Because of the considerable disagreement between these results and others previously obtained, the present writer was led to make two studies along the same line. Before proceeding to present his own data, however, he wishes to offer one comment on Corey's investi-

---

[1] Douglass, Harl R and Peter L. Spencer: Is It Necessary to Weight Exercises in Standard Tests? *Journal of Educational Psychology*, Vol. XIV, February, 1923, pp. 109–112.

[2] Corey, Stephen Maxwell: The Effect of Weighting Exercises in a New Type of Examination. *Journal of Educational Psychology*, Vol. XXI, May, 1930, pp. 383–385.

gation. It does not appear in the article, but the writer happens to know that five of the six instructors assigned weights of zero to some of the elements, and thus altered the conditions of the experiment. A test which contains some items that do not count in determining the score is for all practical purposes no longer the same test as if all the elements were counted, so that the general effect of this factor was to lower the obtained correlations.

Part of the data which the writer wishes to present are based upon a fifty element multiple-answer test with four suggested answers to each element that was taken by sixty-two students in educational measurements. The elements in the test were weighted upon five different bases in addition to the raw score. These bases will be referred to by subscripts as follows:

1 = raw score, one point to each element
2 = percentage of students answering correctly
3 = difficulty as determined through application of normal curve to 2
4 = random distribution of weights one to five
5 = random distribution of weights one to ten
6 = second random distribution of weights one to five

It will readily be seen that the bases of weighting employed in this experiment were such as to give no promise of agreement between the weights, thus making the conditions of the experiment much more unfavorable to securing high correlations between raw and weighted scores than if weights had been assigned according to the opinions of instructors who might be expected to agree to some extent at least. Not only this, but in the case of bases two and three a high negative correlation might be expected, since one was based directly upon how easy the elements were and the other somewhat less directly upon their difficulty. The correlations between the various lists of weights are given in Table I.

TABLE I.—COEFFICIENTS OF CORRELATION BETWEEN WEIGHTS OF ELEMENTS IN FIRST STUDY

$r_{12} = -.62$	$r_{34} = .03$	$r_{45} = -.21$	$r_{56} = .03$
$r_{24} = -.03$	$r_{35} = .24$	$r_{46} = .07$	
$r_{25} = -.14$	$r_{36} = .19$		
$r_{26} = -.11$			

The writer's second study was similar except that instead of the three random assignments of weights, three instructors actually assigned them, and that the weights for ease and difficulty were both taken on the same basis so that the correlation between them was

−1.00. The test in this case was a twenty-two element single-answer test in methods of teaching. The coefficients of correlation between the weights in this case are in Table II.

TABLE II.—COEFFICIENTS OF CORRELATION BETWEEN WEIGHTS OF ELEMENTS IN SECOND STUDY

$r_{23} = -1.00$	$r_{34} = .08$	$r_{45} = .75$	$r_{56} = .34$
$r_{24} = -.09$	$r_{35} = .07$	$r_{46} = .67$	
$r_{25} = -.05$	$r_{36} = -.09$		
$r_{26} = .04$			

It will be seen that except in the case of the weightings according to ease and difficulty and of those by the three instructors, the correlations between weights were practically zero. Those given by the three instructors correlated positively, in one case decidedly low, and in the other two only fairly high.

After weights had been determined as described, each student's score was computed according to each of the six plans of weighting. In the first study there were two ways under each, one by totaling the points allowed for all the elements to which he gave the correct answers, and the other by means of the approved formula for multiple-answer tests, score $= R - \frac{W}{N-1}$, which in this case becomes score $= R - \frac{W}{3}$. In this, of course, R stands for the number of right responses, and W for the number of wrong responses. In the second study only one method, the number correct, was employed. The intercorrelations between the scores on each of the six bases are given in Table III. In the first half thereof the first coefficients are for scores based upon correct responses only, and the second for scores based upon the formula.

TABLE III.—COEFFICIENTS OF CORRELATION BETWEEN SCORES BASED ON DIFFERENT WEIGHTINGS

First Study

$r_{12} = .98, .97$	$r_{23} = .83, .80$	$r_{34} = .90, .88$	$r_{45} = .96, .96$	$r_{56} = .98, .99$
$r_{13} = .92, .92$	$r_{24} = .97, .96$	$r_{35} = .92, .93$	$r_{46} = .98, .98$	
$r_{14} = .98, .98$	$r_{25} = .96, .96$	$r_{36} = .94, .93$		
$r_{15} = .99, .99$	$r_{26} = .95, .95$			
$r_{16} = .99, .99$				

Second Study

$r_{12} = .97$	$r_{23} = .75$	$r_{34} = .94$	$r_{45} = .99$	$r_{56} = .92$
$r_{13} = .95$	$r_{24} = .91$	$r_{35} = .97$	$r_{46} = .98$	
$r_{14} = .98$	$r_{25} = .93$	$r_{36} = .96$		
$r_{15} = .97$	$r_{26} = .96$			
$r_{16} = .98$				

It will be seen that only in the case of the third basis, that is, the weights determined according to the normal probability curve as applied to difficulty, were the coefficients between the raw scores and the weighted scores lower than .97, and that in half of the other cases in the first study they are .99, and in all the others in both studies .97 or .98. Furthermore, among the various bases of unequal weighting the only cases in which the coefficients dropped markedly below .90 were those in which bases two and three were concerned. Such a result would be expected from the fact that there was a fairly high negative correlation between the weights given on these bases. Of the three between the different instructors in the second study, two are decidedly high, .98 and .99, whereas the other one is considerably lower, being only .92.

The comment should probably be inserted that the tendency of the correlations and other measures obtained from the results of the first study to show greater agreement than those from the second is undoubtedly due to the fact that the test employed therein was much longer than that used in the second.

Following Corey's example, the writer also computed the effect of different weights upon marks. This was done in the same way as by Corey, considering the 7 per cent highest scores as *A*'s, the next 24 per cent as *B*'s, the next 38 per cent as *C*'s, the next 24 per cent as *D*'s, and the lowest 7 per cent as *E*'s. Table IV gives figures showing the per cents of the pupils receiving the same letter mark according to the various numbers of bases of weighting.

TABLE IV.—PERCENTAGES OF PAPERS RECEIVING SAME MARKS ACCORDING TO DIFFERENT WEIGHTINGS

Number of different bases of weighting	Percentages given same marks	
	First study	Second study
6	42, 39	38
5	36, 32	33
4	12, 18	15
3	10, 11	14

From this table it appears that about 40 per cent of the students would have received the same letter mark according to all six bases of weighting, and about another 35 per cent according to five of the six. In

other words, in about three-fourths of the cases would either all, or all but one, of the different weightings have given the same mark.

The percentages of cases in which the marks assigned according to the different bases except the first differed from it were also determined. These are shown in Table V.

TABLE V.—PERCENTAGES OF PAPERS ON WHICH MARKS ACCORDING TO DIFFERENT WEIGHTINGS VARIED FROM THOSE BASED ON RAW SCORES

Basis of weighting	Percentage of marks differing from those according to raw scores	
	First study	Second study
2	19, 23	24
3	26, 29	30
4	13, 21	14
5	11, 19	18
6	6, 18	12

A comparison of the results obtained in the writer's two experiments with those of Corey shows that they indicate much less discrepancy or disagreement between scores assigned according to different bases of weighting. In view of the fact that the tests were so short, one containing only fifty elements and the other only twenty-two, the results are even more significant. To the writer they seem to offer very strong support of the conclusion that for new-type tests there is so little to be gained by unequally weighting the elements that it is not worth the labor of assigning such weights and computing scores from them. The longer the test the more fully does this conclusion hold.

THE HANDWRITING OF INDIANS

THOMAS R. GARTH

University of Denver

In order to ascertain whether or not there are racial differences in native traits we shall have to measure those traits. Let us take the simple matter of differences in handwriting between Whites and Indians. It is believed that there are peculiarities in handwriting which run in families in a race. The question is, do these peculiarities run in races? Is Indian handwriting peculiar to itself and different from that of Whites?

Immediately the problem becomes a large one. So many things need to be considered. There is the question of legibility and speed. Another is the question of esthetic quality. Still another is what has been called by graphologists "characterological" interpretation. Yet another is consideration of the diagnosis of the handwriting of the two races. In the present study we have been compelled to confine our inquiry to the first question, *i.e.*, a Comparison of the Legibility and Speed of the Handwriting of Whites and Indians.

Problem

We have asked ourselves these questions;

1. Do these full- and mixed-blood Indians write as legibly as White children of a like school grade?
2. Do they write with the same speed?
3. Which of the factors under consideration have the greater influence on legibility, as age, school grade, degree of White blood?
4. Which of the above factors has the greatest influence on speed?
5. Does this study indicate that the Indian race is mentally retarded?

Materials

The Thorndike Handwriting Scale was used in measuring samples of handwriting of Whites and Indians. The rating was done by students in classes in educational measurements. A sample was often rated by as many as ten individuals, and the average of the ratings taken as the measure. The speed was determined by dividing the total number of letters by the total time of writing. Because the

number of cases of the Whites was so small, we have been compelled to resort to the use of Starch's norms for school grades.

Composition of the Groups and the Results of Measurement

Table I will show the racial and educational composition of the groups. There were six hundred three full-blood Indians and one hun-

Table I

Grade	4	5	6	7	8	Total
Full-bloods	37	64	129	191	182	603
Mixed-bloods	14	35	35	64	42	196
Whites	38	58	61	58	45	260

Table II

Grade	4		5		6		7		8	
	Median	Quartile	Median	Quartile	Median	Quartile	Median	Quartile	Median	Quartile
Group:										
Full-bloods	14.6	2.2	15.8	1.7	15.0	1.6	15.5	1.3	16.6	1.2
Mixed-bloods	16.2	1.2	14.7	1.3	16.8	1.3	16.1	1.2	16.7	.8
Whites	10.1	.7	10.8	.6	12.1	.6	12.6	.5	13.6	.8
Legibility:										
Full-bloods	8.33	.93	8.74	1.01	8.42	.83	9.28	1.12	10.36	1.29
Mixed-bloods	8.33	.96	7.81	.89	8.75	.89	8.81	.95	9.91	1.10
Whites	8.64	.91	8.78	.93	9.88	.87	9.86	1.01	9.90	.83
Starch's norms	8.7	.....	9.3	.....	9.8	.....	10.8	.....	10.9	
Full-bloods SN overlap.	41 per cent	.....	36 per cent	.....	14 per cent	.....	21 per cent	.....	37 per cent	
Mixed-bloods overlap	43 per cent	.....	9 per cent	.....	20 per cent	.....	17 per cent	.....	28 per cent	
Speed:										
Full-bloods	41.67	20.47	53.85	13.22	64.35	12.29	70.78	14.12	69.47	13.49
Mixed-bloods	45.00	12.21	47.72	9.78	57.50	11.46	72.5	13.34	64.28	16.50
Whites	33.08	8.85	46.92	10.57	46.54	11.17	53.50	18.32	68.13	15.39
Starch's norms	47	.....	57	.....	65	.....	75	.....	83	
Full-bloods SN overlap.	46 per cent	.....	44 per cent	.....	48 per cent	.....	43 per cent	.....	19 per cent	
Mixed-bloods overlap	49 per cent	.....	37 per cent	.....	34 per cent	.....	45 per cent	.....	25 per cent	

dred ninety-six mixed-blood Indians found in the United States Indian Schools at Chilocco, Oklahoma, and Albuquerque, New Mexico; and there were two hundred sixty White children taken from city and suburban schools of Colorado. Table II will show the medians of

ages and scores for the several grades along with the measures of overlapping, while Table III will show certain correlations found between the various factors under consideration, *i.e.*, legibility, speed, age, school grade, and degree of blood.

Only when undertaking to find correlations between degree of blood and performance was the mixed-blood group enlarged by some Whites and full-blood Indians taken at random.

TABLE III

	(1) Score	(2) Speed	(3) Age	(4) Grade	(5) Degree
Full-bloods	$r_{12} = .245 \pm .025$		$r_{13} = .004 \pm .027$	$r_{14} = .39 \pm .023$	
			$r_{23} = -.20 \pm .03$	$r_{24} = .32 \pm .024$	
				$r_{34} = .25 = .025$	
Mixed-bloods	$r_{12} = .199 \pm .043$		$r_{13} = .027 \pm .045$	$r_{14} = .42 \pm .036$	$r_{15} = .053 \pm .04$
			$r_{23} = .09 \pm .044$	$r_{24} = .37 \pm .038$	$r_{25} = -.05 \pm .04$
				$r_{34} = .23 \pm .042$	$r_{35} = -.17 \pm .04$
					$r_{45} = .006 \pm .05$
Whites	$r_{12} = .205 \pm .04$		$r_{13} = .284 \pm .038$	$r_{14} = .45 \pm .04$	
			$r_{23} = .44 \pm .03$	$r_{24} = .65 \pm .03$	
				$r_{34} = .92 \pm .006$	

TABLE IV.—AS MULTIPLE CORRELATIONS FOR MIXED-BLOODS, WITH AGE ELIMINATED

I. Score legibility (1), grade (2), degree (3).

$R_1 (23) = .33$ Using $x_1 = b_{12.3}x_2 + b_{13.2}x_3$

$x_1 = .38x2 + .035x_3$

Weight of school grade = .38. Weight of degree of white blood = .036.

II. Score speed (1), grade (2), degree (3). $R_1(23) = .26$

$x_1 = b_{12.3}x_2 + b_{13.2}x_3$. $x_1 = .352x_2 - .057x_3$

Weight of school grade = .352. Weight of degree of White blood = − .057.

INTERPRETATION

It will be seen upon examination of Table II that the Indians, both full- and mixed-bloods, are much older than the Whites. For a school grade the ages of the full-bloods run on the average three years, and the mixed-bloods run four years older than the Whites. The range of age for full-bloods is from ten to twenty-three years, for mixed-bloods is from ten to twenty-three, and for the Whites is from eight to fifteen. However, the median scores of legibility for the racial groups are just about the same, *i.e.*, full-bloods, 9.2; mixed-bloods, 8.8; and Whites 9.4. On the other hand the speed of the full-bloods is greater than that of either the mixed-bloods or the Whites. In turn the speed of the mixed-bloods is greater than that of the Whites—these median scores for speed are respectively 66.9, 57.9, and 50.2 letters per minute. The reader must not conclude

that the difference in speed is due to race as appears on the surface. This will be discussed later on.

It will be better to examine Table II for a comparison of legibility scores and speed scores by grades for the racial groups. It will be seen here that performances of the full-bloods are on the whole more nearly like the White norms from Starch, (as is indicated by the overlapping) than are those of the mixed-bloods. However, they both appear as rather inferior in performance to Starch's norms. They are more nearly like our medians for Whites. That is to say our Whites suffer in comparison with the Starch norms. If race made a difference, the mixed-bloods ought to show up better upon comparison than the full-bloods, but they do not. The same is true for speed. The full-blood performances are more nearly like the Starch norms for speed than are the mixed-bloods as judged by the overlapping.

If the reader will refer to Tables III and IV, he will see that the greatest correlation is found for all groups between legibility and school grade, and speed and school grade with exceptions found in the White data. Age is a small factor excepting in the case of the Whites where it is found to be significantly correlated with speed (.44) and slightly with legibility (.28). For Whites there is the usual high correlation between age and grade (.92). One should notice the insignificant correlation coefficients for degree of blood and Legibility and Speed. They are respectively .053 and −.051. Multiple coefficients of correlation (Table IV) were found between legibility and school grade and degree, and speed, school grade and degree for the mixed-blood group. These are respectively .33 and .26. These are not very high. They were calculated for the purpose of determining the weight that should be assigned respectively to school grade and degree of White blood. For Legibility it will be seen that the ratio for school grade and degree is ten to one, and for speed, about six to one with the one a minus quantity. It will be seen that degree of blood is a matter of small significance. Elsewhere it has been shown that instead of degree of blood having the weight, it seems to have the situation may possibly be due to the influence of social status going with the degree of blood. At any rate degree is of extremely small weight.

Summary

What, then, does this study show?

1. The full-bloods are on the whole somewhat better in legibility than the mixed-bloods, and the Whites are best of all.

2. The Indians are no more nor less speedy than Whites excepting in the Eighth Grade where the Whites excel.

3. The greatest factor in influencing score in legibility and speed for the Indians is school grade. Age has little significance, and degree of White blood practically no significance if any.

4. This study does not indicate that Indians are mentally retarded at least as this is shown in respect to ability in handwriting. It is seen that with proper training Indians may compare well with Whites in handwriting.

A MEASUREMENT OF THE KNOWLEDGE OF PSYCHOLOGY BEFORE AND AFTER FORMAL TRAINING

CALVIN HALL

University of California

The purpose of this investigation was five-fold. We wished to find: (1) The amount of improvement on a comprehensive objective examination in elementary psychology before and after taking a course in that subject; (2) the relation between the scores on the first and second taking of the test; (3) the relation between first and second taking of the test and the score on a college aptitude test; (4) the relation between gains—that is, score on the last examination minus score on the first examination—and college aptitude score; and (5) the effect of training on test reliability.

The first and last examinations were identical and were given on the initial and final meetings of the class. The examination consisted of one hundred eighteen true-false items based upon Perrin and Klein's "Psychology" the chief textbook used. The Washington College Aptitude test was given as a measure of college aptitude. This test is made up of the usual "intelligence" test items. The subjects were all members of a class in Psychology I*A* at the University of California during the summer of 1930. This course extends over a period of six weeks.

The data are presented in Tables I and II.

TABLE I.—SUMMARY OF POPULATIONS, MEANS, SIGMAS, STANDARD ERRORS OF SIGMAS, AND COEFFICIENTS OF VARIATION

Variable	N	Mean	SD	SD of $\sigma_{dist.}$	Coefficient of variation
College aptitude.....	100	121.02	28.328		
First examination.....	94	72.56	5.868	±.428	8.087
Last examination.....	94	92.76	6.928	±.505	7.469

The lack of or negligible correlations indicate that little relationship exists between the variables we were investigating. Considering the corrected coefficients alone, however, certain trends are noticeable. Thus, whatever is measured by the Washington College Aptitude test has more in common with the first presentation of the psychology

examination than with the last. This is to be expected. "Intelligence" should be a more important determinant of information before than after formal training because other factors enter into the final examination, *e.g.*, interest, concentration, habits of study, to lessen the effect of "intelligence." The lack of correlation between "gains" and college aptitude score can be interpreted in two ways. Either improvement in this specific course is not dependent upon any of the factors that determine "intelligence" test score or the law of diminishing returns may operate to enable those doing poorly on the first taking of the

TABLE II.—SUMMARY OF UNCORRECTED AND CORRECTED CORRELATION COEFFICIENTS AND RELIABILITY COEFFICIENTS WITH PROBABLE ERRORS. (UNCORRECTED *r's* IN PLAIN PRINT, CORRECTED *r's* BY UNDERSCORING AND RELIABILITY *r's* IN ITALIC

Variable	College aptitude	First examination	Last examination	Gains
College aptitude	*.975* ± *.003*	**.293**	**.152**	
First examination	.183 ± .068	*.400* ± *.080*	**.484**	
Last examination	.128 ± .066	.261 ± .065	*.730* ± *.042*	
Gains	−.024 ± .069	.....	.....	*.655* ± *.044*

examination to make greater gains than their more informed classmates. The first and last taking of the examination correlate just moderately but probably as well as any two course examinations. Certainly the score on our examination before formal training cannot be used to predict what score will be made on the same examination after training.

The objection may be raised that we did not validate our examination. This objection does not vitally affect our results, however. We assumed that we were measuring knowledge of psychology when we selected questions from the textbook just as every instructor does when he constructs an achievement test for grading purposes.

Further conclusions that should be noted are the following:

1. Although the absolute variability on the last examination increased over that of the first, the difference between the standard deviations is not significant. The relative variability as measured by

the coefficient of variation reveals the opposite tendency. The last examination was ninety-two per cent as variable as the first examination. This difference however is probably not significant.

2. The reliability coefficient of the last examination increased markedly over that of the first examination. This demonstrates the fact that the subjects tested as well as the inherent nature of the test may contribute to the size of the reliability coefficient. Hence any statement of a reliability coefficient for a given test should be accompanied by a description of the subjects.

3. The use of tests for predicting success in specific school courses is of doubtful value if the test is not standardized and possesses a low or unknown reliability. Such tests are often used as pedagogical devices for selecting students where the class is restricted and the writer on the basis of his data condemns this practice.

C. H. STOELTING CO.

MANUFACTURERS—PUBLISHERS—IMPORTERS—EXPORTERS

Apparatus, Tests

AND

Supplies

FOR

PSYCHOLOGY, PSYCHOMETRY, PSYCHOTECHNOLOGY,
PSYCHIATRY, NEUROLOGY, ANTHROPOLOGY,
PHONETICS, PHYSIOLOGY,
AND
PHARMACOLOGY

424 N. Homan Ave., Chicago, Ill., U.S.A.

TESTS AND MEASUREMENTS for Teachers

By ERNEST W. TIEGS, *Dean, University College, University of Southern California.*

Dr. Tiegs here gives those fundamentals of tests and measurements, a knowledge of which is necessary for accomplishing effective results in the use of tests on the elementary, secondary and college levels. The book discusses measurement, the use of tests for definite educational purposes,—the construction, administration, and interpretation of informational objective tests, and the selection and nature of standardized tests. *Riverside Textbooks in Education.* $2.40

DIAGNOSTIC AND REMEDIAL TEACHING

By LEO J. BRUECKNER, *Professor of Elementary Education, University of Minnesota,* and ERNEST O. MELBY, *Associate Professor of Education, Northwestern University.*

Emphasis in this book is placed on the techniques in test and remedial work that experience has proved can be most effectively employed by the average teacher. In addition to the general discussion of the individual pupil and educational tests, there are special chapters on diagnostic and remedial teaching in Arithmetic, Reading, Language, Spelling, Handwriting, the Social Sciences, Character Education, and Health Education. *Riverside Textbooks in Education.* $2.75

Boston New York Chicago **HOUGHTON MIFFLIN CO.** *Dallas Atlanta San Francisco*

Important New Texts

AN INTRODUCTION TO THE SCIENTIFIC STUDY OF CLASSROOM TEACHING

By A. S. Barr

Prof. of Education, University of Wisconsin

A general survey of the scientific study of classroom supervision, covering such subjects as the background of scientific supervision, the objectives of education, the search for the causes of poor work, data-gathering devices, measurement of results, and research. It contains many very pertinent criticisms of current supervisory practices, scientific and otherwise. $2.50

THE ART OF THE TEACHER

By P. F. Valentine

Prof. of Education, San Francisco State Teachers College

In the form of literary essays the author develops a unique approach to a philosophy of education through his conception of teaching as an art. A variety of the most recent points of view in fields related to education is included. $2.00

MAKING HOMES

By Hazel Shultz

Recent Head of Dept. of Home Economics, Univ. of Chicago Laboratory Schools

A valuable home economics text for 9th to 11th year courses such as "Home Making," "The House and Its Care," etc. $2.00

METHODS IN EDUCATIONAL RESEARCH

By Frederick Lamson Whitney

Dir. Educational Research and Prof. of Education, Colorado State Teachers College

A text for courses in "Research and Thesis Work," "Thesis Writing," and "Educational Research." Its objectives are to offer students an opportunity to develop a conception of the scientific method, and to aid the beginner in educational research to attain some skill in gathering, classifying, and interpreting objective material. $2.25

These Are Appleton Books

D. APPLETON AND COMPANY

35 West 32nd Street New York City

By WILLIAM HENRY PYLE

The Psychology of the Common Branches

The book is not only one that can be used as a textbook but it is an exceptionally good reference book. Most of the books of the past ten years on this subject are out of date. Here is one that has brought all of the fundamental subjects up to date. *Journal of Educational Research.*

$2.20 plus 10 cents postage.

WARWICK AND YORK

Publishers BALTIMORE, MD.

100 YEARS OF EXPERIENCE

Back of its Scholarship and Accuracy

That is why The Merriam-Webster is the "Supreme Authority"; is the most widely used reference work in the world; is universally accepted in the courts, colleges, schools, and libraries.

WEBSTER'S NEW INTERNATIONAL DICTIONARY

The "Supreme Authority"

Every State that has adopted an unabridged dictionary for use in its schools has chosen exclusively The Merriam-Webster.

Nearly 100% of all schoolbooks are based on the New International for pronunciation, spelling, compounding, and division of words.

The Colleges voted overwhelmingly in favor of Webster as the standard of pronunciation, in answer to questions submitted by the Chicago Woman's Club.

GET THE BEST!

Write for Free richly illustrated pamphlet with sample pages of The New International.

G. & C. MERRIAM COMPANY

Springfield, Massachusetts

THE MAPLE PRESS COMPANY, YORK, PA.

INCHES
1
2
3
4
5
6
C2399
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
METRIC